PB84-181957

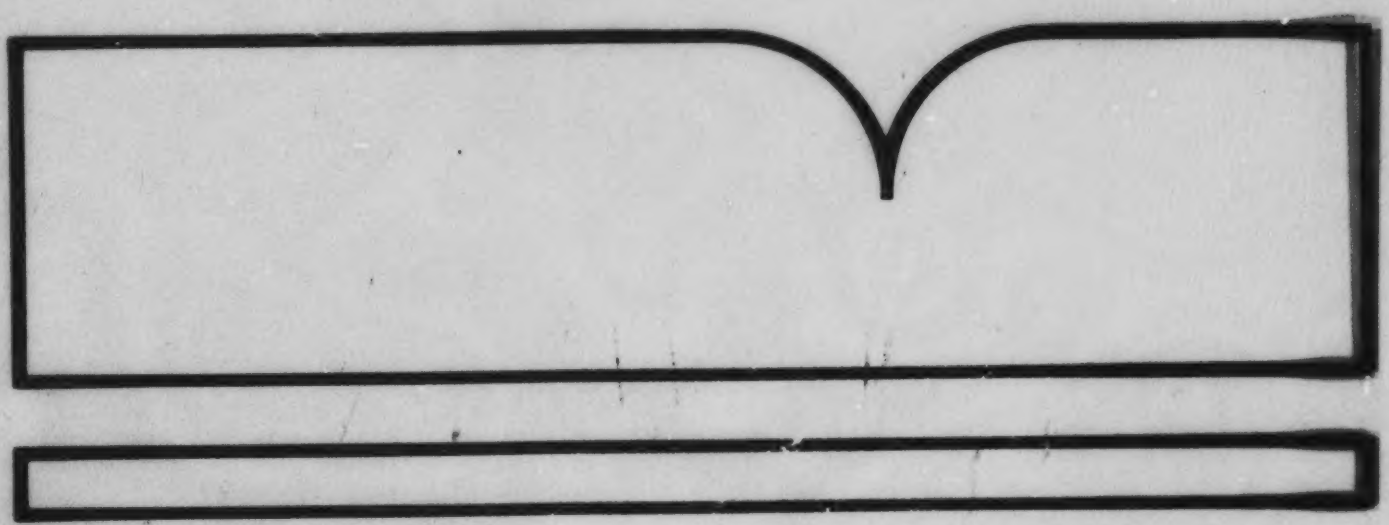
Final Sampling Report for the Study of Personal CO (Carbon Monoxide) Exposure

Research Triangle Inst.
Research Triangle Park, NC

Prepared for

Environmental Monitoring Systems Lab.
Research Triangle Park, NC

Apr 84



U.S. Department of Commerce
National Technical Information Service
NTIS

PB84-181957

EPA-600/4-84-034
April 1984

FINAL SAMPLING REPORT FOR THE STUDY OF PERSONAL CO EXPOSURE

by

Roy W. Whitmore
Shelton M. Jones
Martin S. Rosenzweig

EPA Project Officer
Gerald G. Akland

EPA Contract No.: 68-02-3679

ENVIRONMENTAL MONITORING SYSTEMS LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/4-84-034		PB84 181957

4. TITLE AND SUBTITLE	5. REPORT DATE
Final Sampling Report for the Study of Personal CO Exposure	April 1984
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
R.W. Whitmore, S.M. Jones, M.S. Rosenzweig	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Research Triangle Institute Research Triangle Park, North Carolina 27709	
	11. CONTRACT/GRANT NO.
	68-02-3679
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Environmental Monitoring Systems Laboratory U.S. Environmental Protection Agency (MD-56) Research Triangle Park, North Carolina 27711	Final
	14. SPONSORING AGENCY CODE
	EPA-600/08

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report describes the sample selection procedures used for a study funded by the EPA and conducted by the Research Triangle Institute in 1982 and 1983 to evaluate methodology for collecting data on personal exposure to carbon monoxide (CO). These data were collected in the metropolitan areas of Denver, Colorado and Washington, D.C., using personal exposure monitors (PEMs) and activity diaries. This involved telephone screening of households and selection of household respondents for personal exposure monitoring in the metropolitan areas in and around Denver, Colorado and Washington, D.C. Data on CO breath levels were also collected in Washington, D.C. (PEDCo Environmental conducted the field work in Denver.) The target population in both cities consisted of the non-institutionalized, non-smoking adults (ages 18 to 70) of these metropolitan areas.

Based on the experience gained during the study, the methodology developed, with some modifications, may be used effectively in other areas of the country for collecting PEM data. These modifications should make the methodology more cost effective, improve the response rate, and lead to more accurate activity information.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Personal Exposure Monitoring Carbon Monoxide Exposure Double Sampling for Stratification Telephone Interviewing		

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release to Public	Unclassified	90
	20. SECURITY CLASS (This page)	22. PRICE
	Unclassified	

EPA Form 2220-1 (9-73)

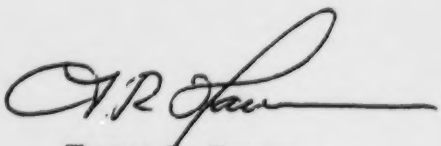
NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

Measurement and monitoring research efforts are designed to anticipate potential environmental problems, to support regulatory actions by developing an in-depth understanding of the nature and processes that impact health and the ecology, to provide innovative means of monitoring compliance with regulations and to evaluate the effectiveness of health and environmental protection efforts through the monitoring of long-term trends. The Environmental Monitoring Systems Laboratory, Research Triangle Park, North Carolina, has the responsibility for: assessment of environmental monitoring technology and systems; implementation of agency-wide quality assurance programs for air pollution measurement systems; and supplying technical support to other groups in the Agency including the Office of Air, Noise and Radiation, the Office of Toxic Substances and the Office of Enforcement.

This document is a report of the sample selection methodology for personal exposure monitoring for carbon monoxide for the residents of Washington, D.C. and Denver, Colorabo during the winter of 1982-1983. This report discusses the sample selection methodology used in the study, the results of applying this methodology, and recommendations for future studies.



Thomas R. Hauser
Director
Environmental Monitoring
Systems Laboratory

TABLE OF CONTENTS

	Page
Disclaimer	ii
Foreword	iii
List of Tables	v
List of Figures	vii
Abstract	viii
1. SAMPLE DESIGN AND WEIGHTING	1
1.1 Introduction	1
1.2 Selection of First-Stage Sampling Units (FSUs)	4
1.3 Selection of Second-Stage Sampling Units (SSUs)	6
1.4 Screening Response and Weight Adjustments	23
1.5 Selection of the Third Stage Sample	33
1.6 Third Stage Response and Sample Weights	40
1.7 Variance Estimation	45
2. SCREENER ANALYSIS	48
2.1 Theoretical Formulation	48
2.2 Results	51
3. METHODOLOGICAL RESULTS	69
3.1 Introduction	69
3.2 Use of Geographically Classified Telephone Directory Listings in Association With Standard Area Household Sampling Techniques	69
3.3 Lead Letter Results	76
3.4 Sampling Person-Days	76
3.5 Conclusions	78
REFERENCES	80

LIST OF TABLES

Table		Page
1.1	Telephone Screening Final Result Codes	12
1.2	Distribution of Initial Telephone Screening Final Result Codes	13
1.3	Stratification of FSUs for the Missed HU Sample	19
1.4	Sample Selection for Washington FSUs with no Donnelley Listings (Montgomery County, Maryland)	22
1.5	Screening Sample Final Result Codes	24
1.6	Distribution of Final Result Codes for Denver Screening Samples	25
1.7	Distribution of Final Result Codes for Washington Screening Samples	26
1.8	Screening Response Rates	27
1.9	Definition of Response Status Variables in Terms of Final Screening Result Codes	29
1.10	Response Distribution for Denver Screening Sample Weighting Classes	31
1.11	Response Distribution for Washington Screening Sample Weighting Classes	32
1.12	Third Stage Sample Allocation for the Denver Sample	35
1.13	Third Stage Sample Allocation for the Washington Sample	36
1.14	Person-Level Weighting Class for the Washington Sample	42
1.15	Distribution of Final Result Codes for Individuals Selected for CO Monitoring in Washington	44
2.1	Estimated Number of Households Using a Fireplace	52
2.2	Estimated Number of Households Using a Wood Stove	52
2.3	Estimated Number of Households Using a Gas Furnace	53
2.4	Estimated Number of Households Using a Gas or Kerosene Space Heater	53
2.5	Estimated Number of Households Using a Gas Cooking Stove	54

LIST OF TABLES (cont.)

Table		Page
2.6	Estimated Number of Households Using a Gas Hot Water Heater	54
2.7	Estimated Number of Households Using a Gas Clothes Dryer	55
2.8	Estimated Number of Households Using Other Gas Appliances.	55
2.9	Estimated Number of Households Having an Attached Garage, or Share a Multi-Family Garage	58
2.10	Estimated Sex Distributions	60
2.11	The Sex Distribution According to the 1980 Census	60
2.12	Estimated Age Distribution - Categorized According to the 1980 Census Definitions	62
2.13	Age Distribution According to the 1980 Census	63
2.14	Estimated Distribution of Relationship to Head of Household	64
2.15	Estimated Distribution of Persons 13 Years and Older Who Smoke or Use Tobacco in Any Form	66
2.16	Estimated Distribution of Persons 13 Years or Older Who Work Either Full or Part Time	67
2.17	Estimated Distribution of Persons 13 Years or Older Who Travel Anywhere at Least 3 Times Per Week	67
2.18	Estimated Distribution of Amount of Time Spent Traveling One Way At Least 3 Times Per Week for Persons 13 Years or Older	68
3.1	Results of Missed HU Checks	71

LIST OF FIGURES

Figure		Page
1.1	Selection of First-Stage Units and the Initial Sample of Donnelley Listings	7
1.2	CO Sample Protocol	14
1.3	Denver CO Sample Protocol	15
1.4	Washington CO Sample Protocol	16

ABSTRACT

This report describes the sample selection procedures used for a study funded by the EPA and conducted by the Research Triangle Institute in 1982 and 1983 to evaluate methodology for collecting data on personal exposure to carbon monoxide (CO). These data were collected in the metropolitan areas of Denver, Colorado and Washington, D.C., using personal exposure monitors (PEMs) and activity diaries. This involved telephone screening of households and selection of household respondents for personal exposure monitoring in the metropolitan areas in and around Denver, Colorado and Washington, D.C. Data on CO breath levels were also collected in Washington, D.C. (PEDCo Environmental conducted the field work in Denver.) The target population in both cities consisted of the non-institutionalized, non-smoking adults (ages 18 to 70) of these metropolitan areas.

Based on the experience gained during the study, the methodology developed, with some modifications, may be used effectively in other areas of the country for collecting PEM data. These modifications should make the methodology more cost effective, improve the response rate, and lead to more accurate activity information.

1. SAMPLE DESIGN AND WEIGHTING

1.1 Introduction

The purpose of the carbon monoxide (CO) study is to monitor personal exposure to carbon monoxide for residents of the metropolitan areas surrounding Denver, Colorado and Washington, D.C. The target populations consist of the non-institutionalized, non-smoking residents, aged 18 to 70, of these metropolitan areas. For the purpose of sample selection, the Washington, D.C. metropolitan area was precisely defined to be all areas simultaneously inside the Washington, D.C. SMSA and inside the Washington, D.C. Urbanized Area as defined by the 1980 Census. The Denver metropolitan area was defined to be the following places in Colorado as defined by the 1980 Census: Denver, Englewood, Arvada, Aurora, and Commerce City. These areas are also in the Urbanized Area of the Denver SMSA.

Among the objectives of the CO study was to make inferences concerning the personal CO exposure for all members of the Denver and Washington metropolitan areas. The only statistically valid procedure that is widely accepted for making such inferences is to select a probability sample from the target populations. Hence, the sample design for the CO study is a deeply stratified, three stage probability-based design. Area sample segments defined by Census geographic variables were selected at the first stage of sampling. Households were selected at the second stage, and all household members were administered a short screening interview. The purpose of this interview was to identify individuals with characteristics believed to be positively correlated with CO exposure so that they could be oversampled in the third stage sample. The third stage sample was a stratified sample of screened eligible individuals. The individuals in the third

stage sample were administered a Computer Model Input Questionnaire and asked to carry a personal CO monitor for 24 or 48 hours for Washington or Denver, respectively. A breath sample was also requested from the individuals who were monitored.

Whenever probability sampling techniques are used in sample surveys sampling weights that reflect the procedures used for sample selection must be used when analyzing the data. The weight of a sample unit can be viewed as the number of units in the target population that the sample unit represents. The initial sampling weight for a unit is usually calculated as the reciprocal of either the probability of selection of the unit or the expected frequency of selection when sampling with replacement. This initial sampling weight is often adjusted in later steps to reduce the bias caused by nonresponse and undercoverage of the target population. The adjusted sampling weights serve to differentially weight the sample data to reflect the level of disproportionality in the final sample relative to the population of interest.

If the sampling weights were all equal, the weights could be ignored in constructing survey estimates. Otherwise, the weights must be used in order to obtain unbiased population estimates. Even when the sample has been designed to affect exact proportional representation of the target population, the differential impact of nonresponse and undercoverage leads to a distortion in the sample which requires the construction of differentially adjusted sample weights. Thus, it would be quite unusual for sample survey data to yield unbiased estimates without the use of sampling weights in the analyses.

Since probability sampling techniques were used to select the individuals to be monitored, the validity of inferences for this survey is

based upon the statistical theory of sample surveys (See, e.g., Cochran [1963] or Raj [1968]). Probability sampling affords unbiased inferences to the target population when sampling weights are used in the analyses.

To the extent that respondents and nonrespondents are alike with respect to probability of responding and/or response values within nonresponse weighting classes, the use of adjusted weights for the analyses reduces the bias due nonresponse. This topic will be discussed further in the sections of this chapter that deal with the nonresponse adjustments for the CO study.

Chapter 2 contains analyses of the screening data, including comparisons to 1980 Census data on age, race, and sex for the target areas. The Census data do not reflect changes in the populations of the target areas between April, 1980, and the Fall of 1982, when the CO study was performed. They do, however, provide a useful benchmark for comparison. The results presented in Chapter 2 show that the sample estimates and Census values are indeed comparable. No such benchmarks are readily available for the persons selected for CO monitoring since these individuals were required to be nonsmokers between 18 and 70 years of age. These benchmark comparisons are not necessary, however, due to the solid probability foundation of the sample selection methodology. The probability structure of the sample provides the basis for valid inferences to the target population.

The purpose of this chapter is to document the sample design and construction of sampling weights for the CO study. Initial, unadjusted sampling weights have been computed for all individuals selected into the monitoring sample. These sampling weights are simply the reciprocal of the overall probability of selection for each individual. Two sets of adjusted

weights have also been computed. The first set of adjusted person-level weights is adjusted only for household-level nonresponse to the screening interviews. The second set of adjusted weights is also adjusted for person-level nonresponse to the CO monitorings. Considerations related to use of these analysis weights are presented in Sections 1.6 and 1.7.

1.2 Selection of First-Stage Sampling Units (FSUs)

The first step in selecting the samples was to extract all block group (BG) and enumeration district (ED) records for each target area from the 1980 Census Summary Tape File 1A (STF-1A) data tapes. In both target areas, all Census records were found to be block group records. However, there is sometimes more than one record for a single block group. Hence, the next step was to produce a data set of records for each site such that there was one and only one record for each block group.

Since some block groups contain few, if any, occupied housing units as reported by the 1980 Census, it was necessary to combine block groups to form a sampling frame of first-stage units (FSUs). The block group records for each site were first ordered by the following Census geographic variables:

State, County, Tract Basic, Tract Suffix, Block Group.

The block groups were then combined within tract to form FSUs with a target size of 40 or more occupied housing units. The BG-level records were processed sequentially and records with a small size measure were combined with the following records until either the last BG record in a tract was reached or the frame unit contained 40 or more occupied housing units. The FSUs were not allowed to cross tract boundaries because of a desire to use sequential selection by tract number to assure geographic dispersion of the sample. In retrospect, the FSUs could have been allowed to cross tract

boundaries without affecting the geographic dispersion of the sample. Each FSU would then contain at least 40 occupied housing units, and undercoverage of tracts with no 1980 Census occupied housing units would be prevented. None of the small frame units were actually selected into the samples, however.

In order to achieve approximately equal probabilities for the second-stage sampling units (SSUs) for each site, the first-stage sampling units were selected with probabilities proportional to size as measured by the 1980 Census counts of occupied housing units. Equal-sized random samples of SSUs within each FSU would then result in approximately equal probabilities for the sample of SSUs.

A sequential, minimum probability replacement (MPR) sampling procedure developed at RTI was used to select the sample of FSUs (See Chromy [1979] and Williams and Chromy [1980]). The frame for each site was first ordered in a serpentine fashion by the following variables:

State, County, Tract, TOTHUS,

where TOTHUS is the number of 1980 Census occupied housing units for the FSU. The FSUs were then selected with probabilities proportional to size as measured by TOTHUS from the ordered sampling frame. The sample sizes were 250 FUSs for Washington and 100 for Denver. Since the sampling was performed with replacement, multiple selections of the same FSU were possible. One FSU in the Denver sample was selected twice so that the Denver sample contains 99 distinct FSUs, or area sample segments. The sample for Washington consists of 250 distinct FSUs or area segments. In addition, the ordering of the sampling frames resulted in a controlled allocation (proportional allocation) to the implicit geographic strata formed by crossing the sorting variables. This control results from the sequential

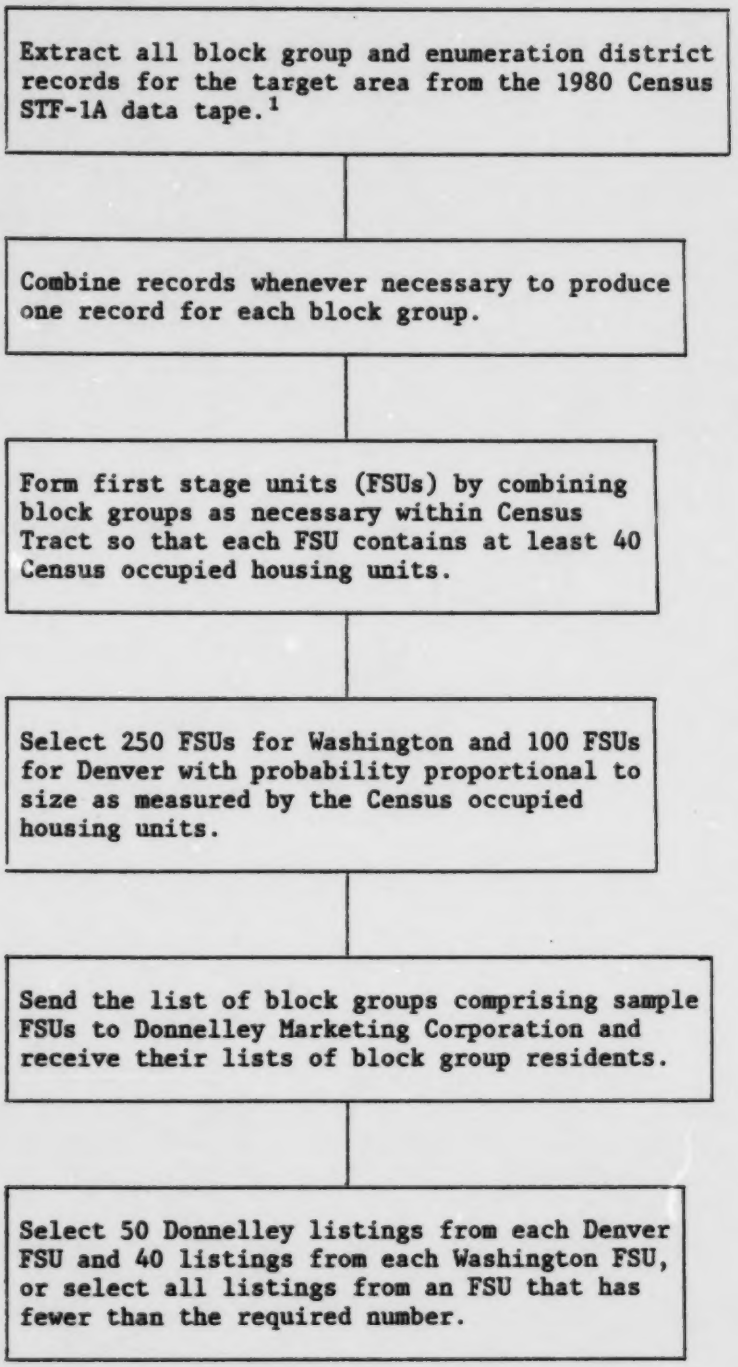
nature of the sample selection and is analogous to selection of a systematic sample from an ordered frame.

After the sample FSUs had been selected, a computer tape listing the block groups (BGs) in the sample FSUs was sent to Donnelley Marketing Corporation. The tape was returned to RTI with name, address, and telephone number listings for each sample BG. The first-stage sample selection and selection of the initial sample of Donnelley listings is summarized by the flow chart in Figure 1.1.

The Donnelley listings had been compiled from two sources: telephone directory listings and vehicle registration records. Ideally, the number of Donnelley listings for a BG should be comparable to the 1980 Census occupied housing units for the BG. The Donnelley list count and the Census count were comparable for most BGs. Unfortunately, however, there were three FSUs in the Washington sample for which there were no Donnelley listings. These three FSUs were in an area of Maryland for which the Donnelley Corporation had no telephone listings, and Maryland does not allow Donnelley Corporation access to its vehicle registration records. For all other FSUs, the Donnelley lists were used as the second-stage sampling frames.

1.3 Selection of Second-Stage Sampling Units (SSUs)

A second-stage sample of housing units, as defined for the 1980 Census, was selected within each sample FSU. For the FSUs with Donnelley listings, the Donnelley listings were used to select the second-stage sample. The listings were used as a second-stage sampling frame in lieu of the traditional field listing of all housing units in the sample segments. For the three Washington FSUs with no Donnelley listings, field procedures were used to select second-stage samples of housing units.


Extract all block group and enumeration district records for the target area from the 1980 Census STF-1A data tape.[1]
Combine records whenever necessary to produce one record for each block group.
Form first stage units (FSUs) by combining block groups as necessary within Census Tract so that each FSU contains at least 40 Census occupied housing units.
Select 250 FSUs for Washington and 100 FSUs for Denver with probability proportional to size as measured by the Census occupied housing units.
Send the list of block groups comprising sample FSUs to Donnelley Marketing Corporation and receive their lists of block group residents.
Select 50 Donnelley listings from each Denver FSU and 40 listings from each Washington FSU, or select all listings from an FSU that has fewer than the required number.

[1]All extracted records were block group records.

Figure 1.1 Selection of First-Stage Units and the Initial Sample of Donnelley Listings

1.3.1 Selection of SSUs within FSUs with Donnelley Listings

For each FSU with Donnelley listings, a simple random sample of listings was selected without replacement. A random sample of 50 listings was selected within each Denver FSU, and 40 within each Washington FSU. For any FSU that did not contain this many listings, all listings were selected. A sample of 9,876 listings was selected for Washington, and 4,987 listings were selected for the Denver sample. When it became apparent that a smaller sample size would produce sufficient screening data for Washington, the Washington sample was randomly subsampled. The ultimate sample size for Washington was 8,643 listings.

The choice of a fixed number of sample listings from each sample FSU, 40 for Washington and 50 for Denver, was designed to produce an approximately equal probability sample of Donnelley listings. Let S(i) represent the size measure, the 1980 Census count of occupied housing units, for the i-th FSU. Then the first-stage sampling weight for the i-th FSU is simply

$$W_1(i) = \begin{cases} n_i * (1/100) \sum_j S(j) / S(i) \text{ for Denver,} \\ n_i * (1/250) \sum_j S(j) / S(i) \text{ for Washington,} \end{cases} \quad (1)$$

where the summation is over all FSUs on the sampling frame for the particular site, Denver or Washington, and n_i is the number of hits for the i-th FSU.

Let D(i) represent the number of Donnelley listings received for the i-th FSU. If M(i) represents the number of listings selected from the i-th FSU, then the second-stage sampling weight component is

$$W_2(i) = D(i) / M(i). \quad (2)$$

The product of these two weight components is the sampling weight associated with each listing in this initial sample. Hence, the sampling weight

weight for each sample Donnelley listing in the i-th FSU is given by

$$W_1(i) * W_2(i) \ . \qquad (3)$$

When the number of Donnelley listings, D(i), and the Census count of occupied housing units, S(i), are comparable, they essentially cancel each other in equation (3), and the sampling weight becomes approximately

$$W_1(i) * W_2(i) \sim \begin{cases} (1/100) \sum_j S(j)/50 \text{ for Denver,} \\ (1/250) \sum_j S(j)/40 \text{ for Washington.} \end{cases} \qquad (4)$$

The actual sampling weights are unequal to the extent that S(i) and D(i) differ and to the extent that the segment sample sizes, M(i), are unequal.

In order to obtain complete coverage of the target population, the sample of Donnelley listings was regarded as a sample of housing units identified by the name and address in each Donnelley listing. There were many Donnelley listings for which the address alone was not sufficient to identify a specific housing unit. This was particularly true of housing units in apartment complexes. Since most Donnelley listings come from telephone directory listings, the address shown for apartment residents was often no more than the street address of the apartment complex. Whenever the address was not sufficient to identify a particular housing unit, the individual's name was also used. The sample housing unit was then defined to be the housing unit at the sample address in which the named individual either currently or previously resided.

It is also possible for a single housing unit to be linked to more than one Donnelley listing. This can occur whenever the Donnelley Corporation listed multiple telephone subscribers and/or vehicle registrants for a single housing unit. These frame multiplicities can be associated with current residents of the housing unit and/or previous residents of the

housing unit. Since the name and address must sometimes be used simultaneously to identify a sample housing unit, it is virtually impossible to accurately identify all of these frame multiplicities. Consequently, no attempt was made to remove these multiplicities from the sampling frame prior to sample selection. It also does not appear to be cost-effective to pursue any multiplicity adjustments in analysis of this sample. Hence, the sampling weights have been computed treating each sample housing unit as if it was linked to only one Donnelley listing.

All sample Donnelley listings were initially assigned to the telephone mode for administration of the screening interview. About 75 percent of the sampled Denver listings and about 88 percent of the sampled Washington listings had phone numbers. The difference seems to be that vehicle registration records were not available to Donnelley Corporation for the Maryland portion of the Washington sample, but were available for all of Colorado. RTI's telephone interviewing staff attempted to obtain telephone numbers for the sample listings with no phone number. Phone numbers were obtained for about ten percent of these listings.

For each call made by the telephone interviewing staff, the telephone number was first verified with the individual who answered. If the correct telephone number had been reached, the address was also verified. Since the sample was regarded as an address sample, not a telephone number sample, the interview was terminated if the individual's address was not the address shown for the sample Donnelley listing. These addresses were placed in a pool to be subsampled for field interviewing. Otherwise, all individuals living in the housing unit (1980 Census definition) were screened by a telephone interview. The name shown for the Donnelley listing was never verified. It was implicitly assumed that the correct

housing unit had been accessed if the telephone number and address were correct.

Table 1.1 presents the final result codes generated by the telephone screening attempts. It also identifies a pool of result codes for which the listings were treated as not covered by the telephone screenings, e.g., those for which no phone number could be obtained. Some assumptions are inherent in this categorization. For example, it has been assumed that listings with a final result of "ring-no-answer" are listings for which the address would have been verified as correct if someone had been contacted. The distribution of the telephone result codes is presented in Table 1.2.

It was not feasible with the project's time and money constraints to attempt field screenings for all sample Donnelley listings not covered by the telephone screenings. Hence, a subsample of these listings was selected as illustrated by the subsample of n_2 Donnelley listings shown in Figure 1.2. Given the sampling rates shown in Figures 1.3 and 1.4, the loss in precision due to this subsampling is considered acceptable. A field interviewer was sent to each housing unit represented by this subsample of Donnelley listings. The field interviewer administered the screening interview if a reliable respondent was available on the first attempt. Otherwise, the interviewer attempted to get the correct telephone number from a neighbor or information operator. When the field interviewer was able to get a telephone number for the sample housing unit, subsequent screening attempts were made by telephone. When the first attempt did not yield either a completed interview or a telephone number, additional field attempts were made for the Denver sample. The wide geographic dispersion of the Washington sample made additional field attempts prohibitively expensive for the Washington sample.

Table 1.1 Telephone Screening Final Result Codes

Code	Interpretation
01	Completed Interview
51	Refusal or Breakoff, Conversion Attempt Failed
52	Refusal or Breakoff, Not Reached for Conversion Attempt
53[1]	Circumstantial Non-Interview
64[2]	Partial Interview
71[3]	Nonworking Number
72	Nonresidential Number
73	Entire Household Moving
74[3]	Address doesn't match Donnelley Listing
75	No Reliable Respondent (3 Attempts)
76[3]	No Phone Number
81[3,4]	Final Phone Problem (Confirmed by Operator)
82	Ring-No-Answer (6 Attempts)
83	Answering Machine (6 Attempts)
84	Final Busy (10 Attempts)
95	Other (Both Eligible and Ineligible Cases)

[1]An eligible housing unit was contacted, but circumstances prevented a screening interview.

[2]Breakoff after Question 7.

[3]Result Codes for the pool of Donnelley listings not covered by the initial telephone screening.

[4]Branching, complete silence, fast-busy or other problem confirmed by an Operator.

Table 1.2. Distribution of Initial Telephone Screening Final Result Codes

Result Code	Frequency	
	Denver	Washington
01	1997	4245
51	469	673
52	135	93
53	6	10
64	13	32
71	436	997
72	91	187
73	92	187
74	350	609
75	21	36
76	1032	782
81	7	8
82	207	503
83	22	35
84	4	4
95	105	242
Total	4,987	8,643

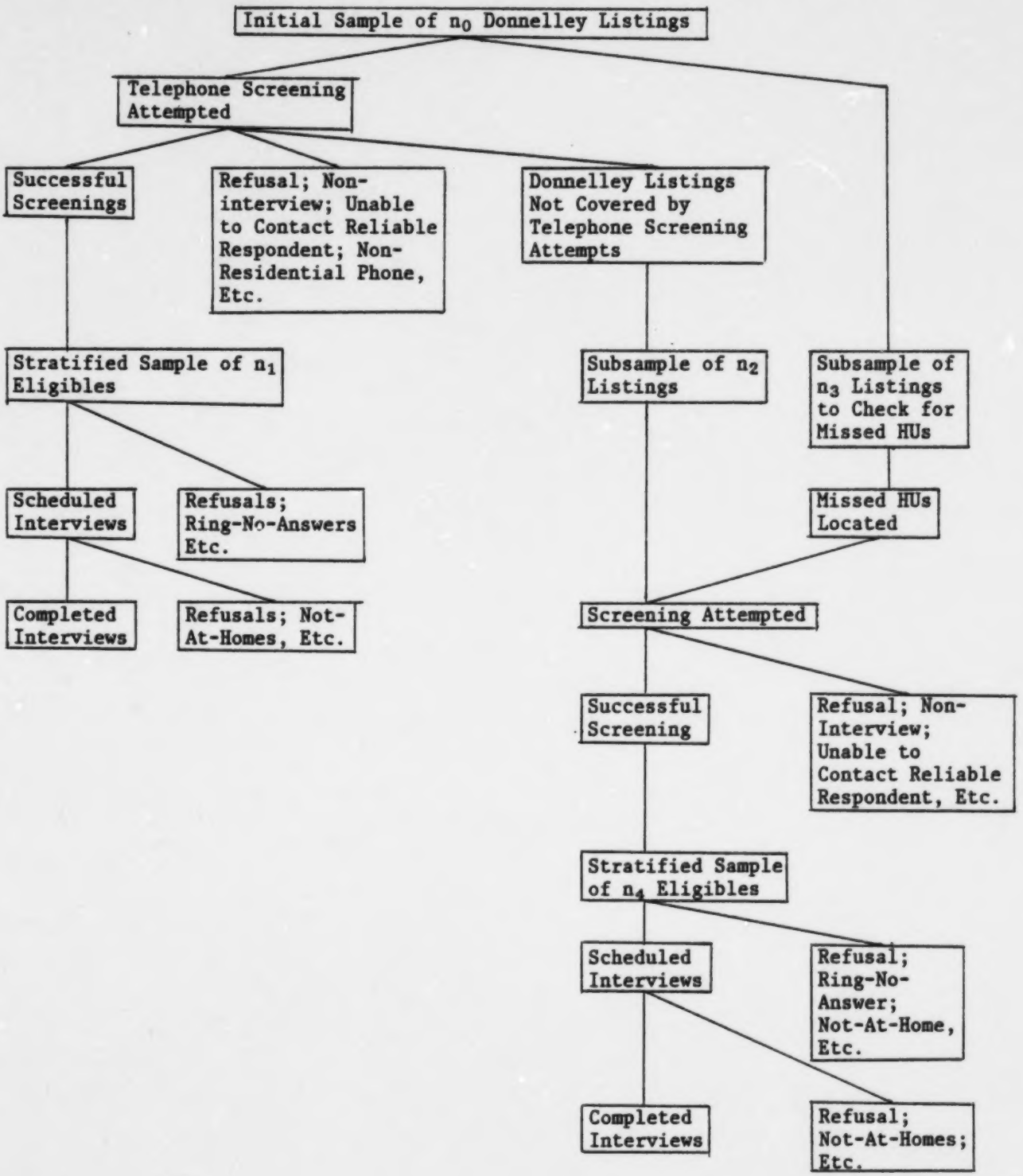

Initial Sample of n_0 Donnelley Listings
Telephone Screening Attempted
Successful Screenings
Refusal; Non-interview; Unable to Contact Reliable Respondent; Non-Residential Phone, Etc.
Donnelley Listings Not Covered by Telephone Screening Attempts
Stratified Sample of n_1 Eligibles
Subsample of n_2 Listings
Subsample of n_3 Listings to Check for Missed HUs
Scheduled Interviews
Refusals; Ring-No-Answers Etc.
Missed HUs Located
Completed Interviews
Refusals; Not-At-Homes, Etc.
Screening Attempted
Successful Screening
Refusal; Non-Interview; Unable to Contact Reliable Respondent, Etc.
Stratified Sample of n_4 Eligibles
Scheduled Interviews
Refusal; Ring-No-Answer; Not-At-Home, Etc.
Completed Interviews
Refusal; Not-At-Homes; Etc.

Figure 1.2. CO SAMPLE PROTOCOL

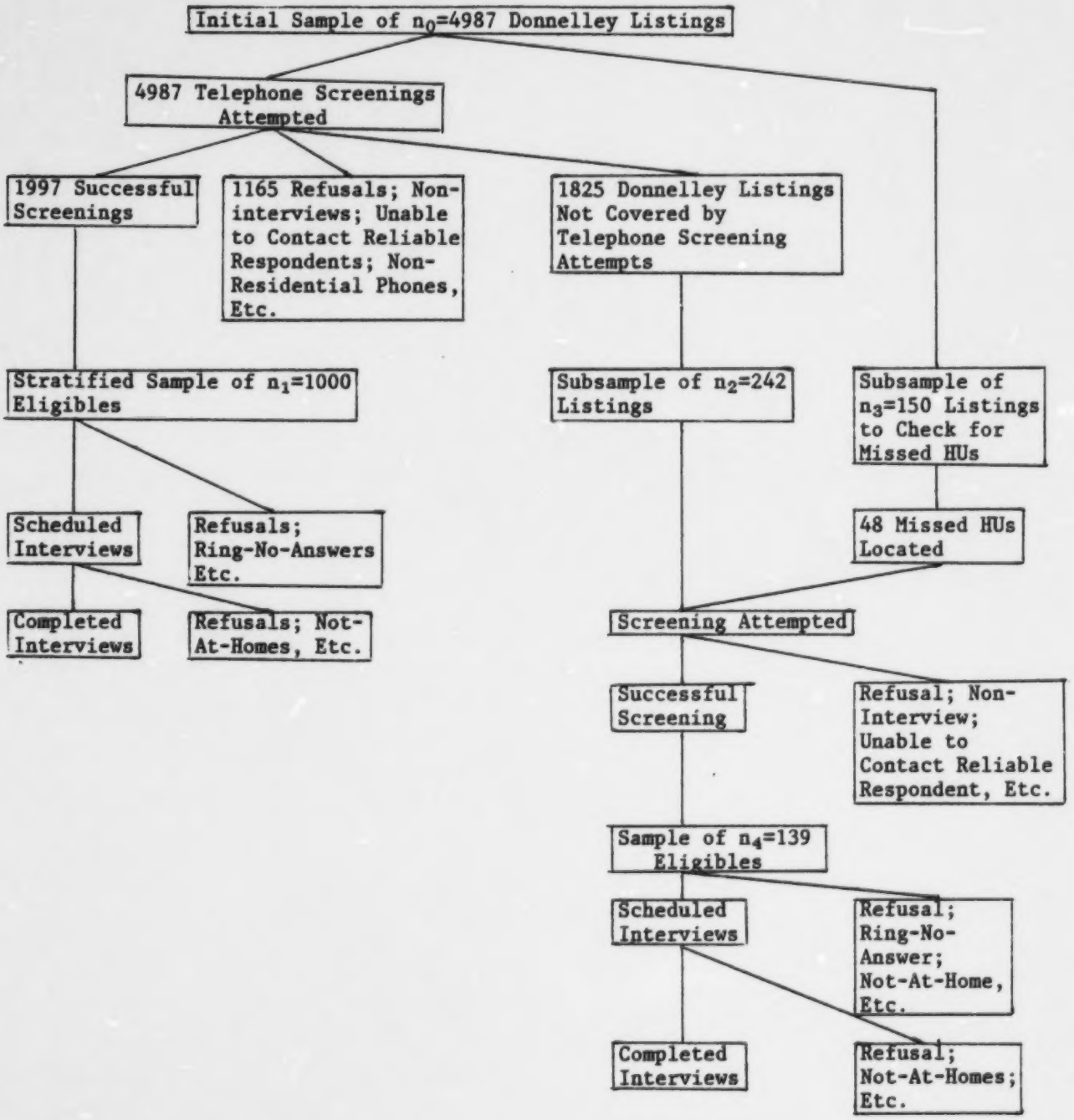
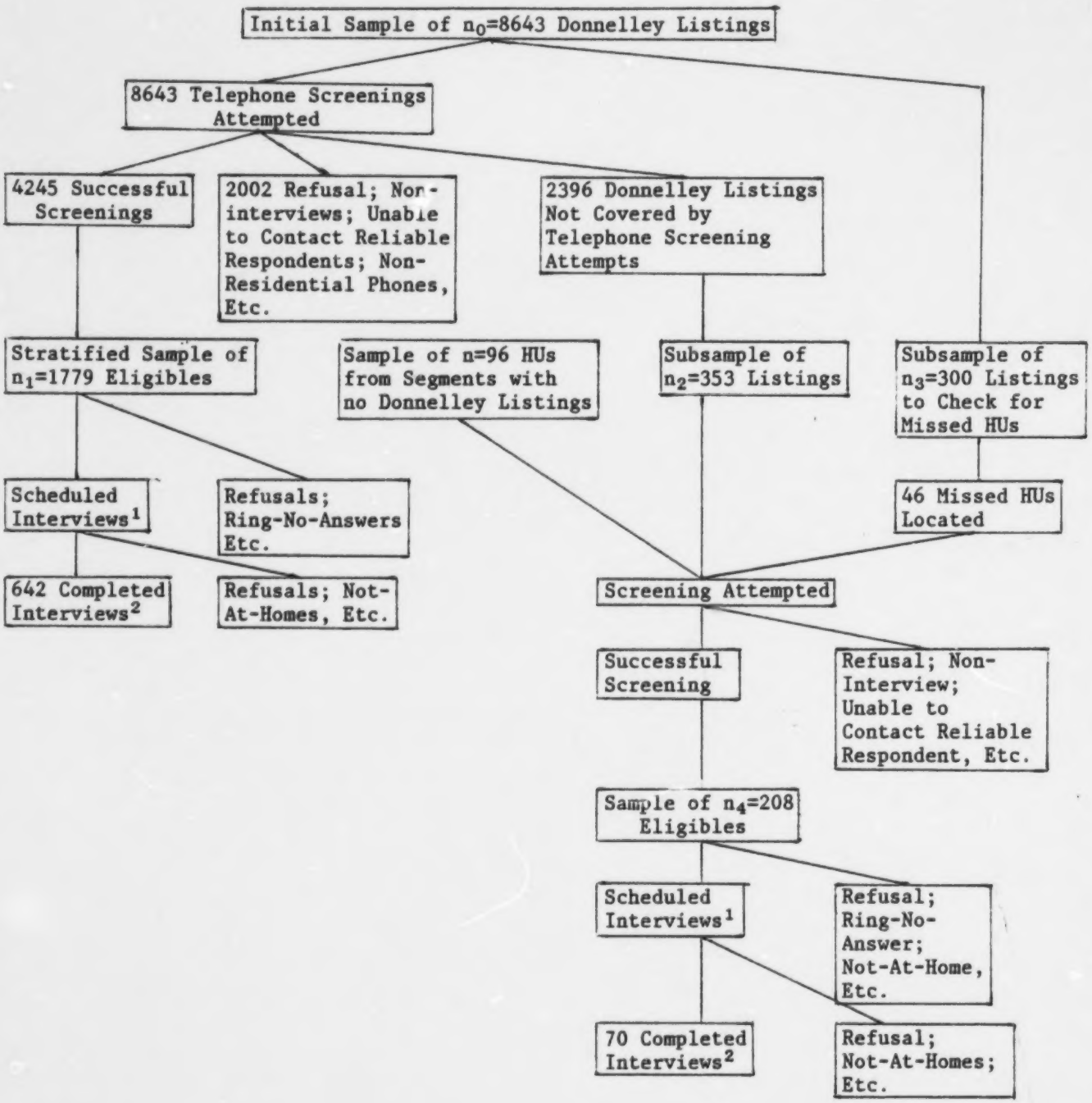

Initial Sample of n_0=4987 Donnelley Listings
4987 Telephone Screenings Attempted
1997 Successful Screenings
1165 Refusals; Non-interviews; Unable to Contact Reliable Respondents; Non-Residential Phones, Etc.
1825 Donnelley Listings Not Covered by Telephone Screening Attempts
Stratified Sample of n_1=1000 Eligibles
Subsample of n_2=242 Listings
Subsample of n_3=150 Listings to Check for Missed HUs
Scheduled Interviews
Refusals; Ring-No-Answers Etc.
48 Missed HUs Located
Completed Interviews
Refusals; Not-At-Homes, Etc.
Screening Attempted
Successful Screening
Refusal; Non-Interview; Unable to Contact Reliable Respondent, Etc.
Sample of n_4=139 Eligibles
Scheduled Interviews
Refusal; Ring-No-Answer; Not-At-Home, Etc.
Completed Interviews
Refusal; Not-At-Homes; Etc.

Figure 1.3. DENVER CO SAMPLE PROTOCOL


Initial Sample of n_0=8643 Donnelley Listings
8643 Telephone Screenings Attempted
4245 Successful Screenings
2002 Refusal; Non-interviews; Unable to Contact Reliable Respondents; Non-Residential Phones, Etc.
2396 Donnelley Listings Not Covered by Telephone Screening Attempts
Stratified Sample of n_1=1779 Eligibles
Sample of n=96 HUs from Segments with no Donnelley Listings
Subsample of n_2=353 Listings
Subsample of n_3=300 Listings to Check for Missed HUs
Scheduled Interviews[1]
Refusals; Ring-No-Answers Etc.
46 Missed HUs Located
642 Completed Interviews[2]
Refusals; Not-At-Homes, Etc.
Screening Attempted
Successful Screening
Refusal; Non-Interview; Unable to Contact Reliable Respondent, Etc.
Sample of n_4=208 Eligibles
Scheduled Interviews[1]
Refusal; Ring-No-Answer; Not-At-Home, Etc.
70 Completed Interviews[2]
Refusal; Not-At-Homes; Etc.

[1]A total of 1217 interviews were scheduled between all Washington samples.

[2]Usable CO monitor data were obtained for 712 individuals in the Washington samples.

Figure 1.4. WASHINGTON CO SAMPLE PROTOCOL

As shown in Figure 1.3, a total of 1,825 sample Donnelley listings were not covered by the initial telephone screenings for Denver. A simple random sample of 242 of these listings was selected without replacement. A field screening attempt was made for each listing in this subsample. Similarly, as shown in Figure 1.4, a simple random sample of 353 listings was selected without replacement from 2,396 sample listings not covered by the initial telephone screenings for Washington. These samples will be referred to henceforth as the "No Previous Contact" samples. A field screening was attempted for each listing in these samples. Since a simple random sample was selected in each case, the sampling weight component is

$$W_3 = \begin{cases} 1825/242 & \text{for Denver,} \\ 2396/353 & \text{for Washington.} \end{cases} \qquad (5)$$

Using equation (3), the sampling weight associated with each Donnelley listing from the i-th FSU in the No Previous Contact sample can be expressed as

$$W_1(i) * W_2(i) * W_3. \qquad (6)$$

The second-stage sample of housing units also addressed the problem of undercoverage by the frame of Donnelley listings. In a traditional listing of all HUs in the sample segments, the interval from each sample housing unit (HU) to the next listed HU is checked for missed HUs. If unlisted HUs are found, they are included in the sample. The direct extension of this procedure to the sample of Donnelley listings would require a check for missed HUs for each of the 8,643 Washington sample HUs and each of the 4,987 Denver sample HUs. This procedure was not feasible since most of the screening interviews were conducted by telephone. Instead, a subsample of Donnelley listings at which to check for missed HUs was selected for each

site. The subsample of n_3 HUs, using the notation from Figure 1.1, was selected in two stages. First, a subsample of FSUs was selected, 15 for Denver and 30 for Washington. Then a subsample of ten Donnelley listings from the initial sample of n_0 Donnelley listings was selected by simple random sampling without replacement from each of these area segments. The FSUs were selected as a stratified simple random sample without replacement with strata as defined in Table 1.3. This stratification was designed to guarantee that some of the missed HU sample would fall within the segments for which the number of 1980 Census occupied housing units was 50 percent or more greater than the number of Donnelley listings. The sampling weight component associated with the selection of FSUs for the Missed HU Sample is easily computed. It is the ratio of the number of FSUs belonging to a stratum divided by the number of FSUs selected from the stratum as shown by $W_4(i)$ in Table 1.3. Similarly, the weight component for the subsample of ten listings from the i-th FSU is

$$W_5(i) = M(i)/10, \tag{7}$$

where M(i) is the number of Donnelley listings from the i-th FSU that were selected into the initial sample of n_0 listings. Hence, the final sampling weight for each of the listings selected into the Missed HU Sample is

$$W_1(i) * W_2(i) * W_4(i) * W_5(i).^1 \tag{8}$$

This sampling weight is then associated with all missed HUs that enter the sample through the i-th FSU.

---

[1]The missed HU sample was selected as a subsample of the initial sample of 9,876 listings for Washington, not the ultimate sample of 8,643 listings. Hence, the value used for M(i) in all factors of this weight was based upon the initial sample.

Table 1.3. Stratification of FSUs for the Missed HU Sample

Site	Stratum Number	Stratum Definition	Total FSUs	Sample FSUs	Sampling Weight, $W_4(i)$
Denver	1	$0 < EHUS^1 \leq 1.5$	92	13	92/13
Denver	2	$EHUS > 1.5$	7	2	7/2
Washington	1	$0 < EHUS \leq 1.5$	204	24	204/24
Washington	2	$1.5 < EHUS \leq 2.5$	34	4	34/4
Washington	3	$EHUS > 2.5$	9	2	9/2

[1]EHUS is defined to be the ratio of the number of 1980 Census occupied housing units for the FSU divided by the number of Donnelley listings for the FSU.

The sample of Donnelley listings for which a missed HU check was performed is considered to be minimally adequate. A larger sample was not selected because the expense of the missed HU checking depended upon the unknown quality of the Donnelley listings. The missed HU field procedure was designed to also detect and document misclassifications of HUs into block groups. Hence, the Missed HU Sample was important for assessing the usefulness of the Donnelley listings as a sampling frame. As a result, it was decided that a thorough check in a few FSUs was the best approach. The results of these checks were generally favorable and are discussed in Chapter 3.

1.3.2 Selection of SSUs within FSUs with no Donnelley Listings

As discussed earlier, there were three FSUs in the Washington sample for which there were no Donnelley listings. Sample housing units were selected from these FSUs even though no funds were budgeted for field listing of housing units in area segments. Hence, a minimal cost procedure that affected complete coverage of these FSUs was adopted. Block statistics from Summary Tape File 1B (STF-1B) were used to obtain the 1980 Census count of occupied housing units for each block within these FSUs. The blocks were then combined into subsegments and one subsegment was selected from within each FSU. The subsegments were selected with probabilties proportional to their 1980 Census count of occupied housing units. Thus, approximately equal sized samples within the subsegments were selected to obtain an approximately equally weighted sample.

Since the budget for field screenings was very limited, the field samples were selected without listing all housing units. Instead, every k-th HU was selected into the sample, with a random start from 1 to k. The value of k was chosen to yield approximately 35 sample housing units in

each subsegment based upon the 1980 Census data. The only compromise resulting from this procedure is that direct quality control of the listing became impractical. However, as an alternative, these samples were listed by one of RTI's most experienced and reliable field staff members. We feel that the quality of the listing was probably superior to the quality usually achieved with the standard procedures.

If a reliable respondent was available on the first pass through each subsegment, a screening interview was conducted. If the field interviewer could get the home telephone number, subsequent screening attempts were made by telephone from RTI. Otherwise, the budget permitted no further interview attempts.

Selection of a sample of 96 HUs from the three Washington segments with no Donnelley listings is documented in Table 1.4. This sample was selected in two stages. First a subsegment was selected from within each segment. Then a sample of HUs was selected from each subsegment. Hence, two sampling weight components are needed. The subsegments were selected with probabilities proportional to size as measured by the 1980 Census count of occupied housing units. Hence, the weight component associated with selection of a subsegment from the i-th FSU is

$$W_6(i) = S(i) \;/\; S(i,\ell), \tag{9}$$

where $S(i)$ is the size of the i-th FSU, as defined before, and $S(i,\ell)$ is the size of the ℓ-th subsegment defined within the i-th FSU, i.e., the number of Census occupied HUs in the subsegment. Hence, the sampling weight for each sample subsegment with no Donnelley listings is

$$W_1(i) * W_6(i) = (1/250) \sum_j S(j) \;/\; S(i,\ell), \tag{10}$$

where $W_1(i)$ is defined by equation (1). Since every k-th HU was selected within each subsegment, the sampling weight for this sampling stage is the

Table 1.4. Sample Selection for Washington FSUs with no Donnelley Listings (Montgomery County, Maryland)

FSU Description		Subsegment Description		Subsegment Sampling	Sampling	Sample	HU Sampling
Tract	Block Group	Block(s)	Seg. #	Weight, $W_6(i)$	Rate, k	Size	Weight, $W_7(i)$
7006.03	1	109,110,110	248	480/210	4	27	4
7013.04	2	201,202	249	343/34	1	34	1
7013.05	2	205	250	1119/100	7	35	14[1]

[1]A sampling rate of k=7 was used for segment 250, i.e., every 7-th HU was listed. However, there was much new construction in this segment and 69 HUs were listed. Since the target sample size for this segment was 30 HUs, a systematic subsample of half of these HUs was selected to be fielded.

integer k itself. However, taking every 7-th HU in segment 250 produced 69 sample HUs, more than twice the target sample size for this segment. Hence, a systematic subsample of half the listed HUs in this segment were selected yielding a weight component of 14 rather than 7. Letting $W_7(i)$ denote the HU sampling weight component, as shown in Table 1.4, the sampling weight for each HU selected from the i-th FSU with no Donnelley listings is computed as

$$W_1(i) * W_6(i) * W_7(i), \qquad (11)$$

applying the segment weight from equation (10).

1.4 Screening Response and Weight Adjustments

The expanded set of final result codes shown in Table 1.5 was used for the field screenings. The distribution of these final result codes for Denver and Washington screening samples is shown in Tables 1.6 and 1.7. The overall response rates for the screening phase, shown in Table 1.8, were 70.4 percent for Denver and 75.8 percent for Washington. The sample design proposed in Chapter 3 is expected to raise these response rates to near 80 percent for future studies of this type.

Initial sampling weights were defined for all sampling units as discussed in the previous sections:

Weight Formula	Sample
Equation (3)	Initial Telephone Screenings
Equation (6)	No Previous Contact Sample
Equation (8)	Missed HUs
Equation (11)	Segments with no Donnelley listings

For each sample unit, the initial screening weight computed as shown above will be referred to as WSCR(i,j) for the j-th housing unit in the i-th sample segment, or the i-th first stage sampling unit. These initial

Table 1.5. Screening Sample Final Result Codes

Code	Interpretation
01	Completed Interview/Telephone Number Obtained
02	Completed Interview/Telephone Number Not Obtained
03	Completed Interview/No Telephone Available
05	Entire Household Moving Based on Field Contact
11	No Age Eligible During Field Contact/Telephone Number Obtained
14	No One Home During Field Contact/Telephone Number Not Obtained
17	Language Barrier/Telephone Number Obtained
18	Language Barrier/Telephone Number Not Obtained
20	Field Refusal/Breakoff
25	No Contact
30	Not a Residence
31	Vacant
32	Demolished/Condemned
33	No Such Address
40	Other Field Result
51	Telephone Refusal or Breakoff/Conversion Attempt Failed
52	Telephone Refusal or Breakoff/No Conversion Attempt
53	Circumstantial Non-Interview
64	Partial Interview
71	Nonworking Number
72	Nonresidential Number
73	Entire Household Moving Based on Telephone Contact
74	Address Doesn't Match Donnelley Listing
75	No Reliable Respondent (3 Attempts)
82	Ring-No-Answer (6 Attempts)
83	Answering Matching (6 Attempts)
84	Final Busy (10 Attempts)
95	Other Telephone Result

Table 1.6. Distribution of Final Result Codes for Denver Screening Samples

Result Code	Primary Phone Screenings	No Previous Contact	Missed HUs	Total
01	1997	109	27	2133
05	0	6	0	6
17	0	4	1	5
18	0	1	0	1
20	0	22	8	30
25	0	29	12	41
30	0	8	0	8
31	0	6	0	6
32	0	3	0	3
33	0	11	0	11
40	0	43	0	43
51	469	0	0	469
52	135	0	0	135
53	6	0	0	6
64	13	0	0	13
72	91	0	0	91
73	92	0	0	92
75	21	0	0	21
82	207	0	0	207
83	22	0	0	22
84	4	0	0	4
95	105	0	0	105
Total	3162	242	48	3452

Table 1.7. Distribution of Final Result Codes for Washington Screening Samples

Result Code	Primary Phone Screenings	No Previous Contact	Missed HUs	No Donnelley Listings	Total
01	4245	85	15	56	4401
02	0	2	1	2	5
03	0	0	0	2	2
05	0	2	0	0	2
11	0	0	1	0	1
14	0	9	1	0	10
18	0	1	0	0	1
20	0	29	3	10	42
25	0	49	17	15	81
30	0	8	0	0	8
31	0	19	0	1	20
32	0	5	0	0	5
33	0	7	0	0	7
40	0	38	6	0	44
51	673	0	0	1	674
52	93	0	0	0	93
53	10	0	0	0	10
64	32	3	0	5	40
71	0	5	0	0	5
72	187	1	0	0	188
73	187	0	1	0	188
74	0	27	0	3	30
75	36	1	0	0	37
82	503	2	1	0	506
83	35	1	0	0	36
84	4	0	0	0	4
95	242	59	0	1	302
Total	6247	353	46	96	6742

Table 1.8. Screening Response Rates[1]

Sample	Denver	Washington
Initial Telephone Screenings	0.720	0.791
No Previous Contact Sample	0.524	0.282[2]
Missed HUs	0.563	0.364[2]
No Donnelley Listings	NA	0.632[2]
Overall	0.704	0.758

[1] Response rate calculation in terms of final result codes from Table 1.7:

$$\text{Rate} = \frac{01 + 02 + 03}{\text{Total} - 05\text{-}30\text{-}31\text{-}32\text{-}33\text{-}72\text{-}73\text{-}82}$$

[2] Field screening response rates are low for Washington because the field effort was minimized to control costs. Due to the wide geographic dispersion of the Washington field sample, a more exhaustive field effort was considered to be prohibitively expensive.

screening weights were then adjusted to reduce the bias due to nonresponse as described in the remainder of this section.

In order to compute the nonresponse adjustments, it was necessary to create indicators of response and eligibility status for each sample housing unit (and/or Donnelley listing). These indicator variables were defined for the j-th housing unit within the i-th first stage sampling unit as follows:

$$RESP(i,j) = \begin{cases} 1 & \text{if the unit responded,} \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

$$ELIG(i,j) = \begin{cases} 1 & \text{if the unit was known to be survey eligible,} \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

$$ELGKWN(i,j) = \begin{cases} 1 & \text{if the eligibility status of the unit was known,} \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

These eligibility and response indicators were based upon the final result codes for the attempted screening interviews as shown in Table 1.9.

The first step in the nonresponse adjustment procedure was to define weighting classes. A primary consideration for formation of weighting classes is that responses should be more homogeneous within classes than between classes. Weighting classes should be homogeneous with respect to response rate and/or response values. Weighting classes are usually based upon design characteristics because one must be able to identify the appropriate class for both respondents and nonrespondents. Each weighting class should contain approximately 20 or more respondents, because the weight adjustments are based upon estimation of the characteristics of each weighting class from its respondents.

Table 1.9. Definition of Response Status Variables in Terms of Final Screening Result Codes

RESP	ELIG	ELGKWN	Final Result Codes
1	1	1	01, 02, 03
0	1	1	20, 51, 52, 53, 64
0	0	1	05, 30, 31, 32, 33, 72, 73
0	0	0	11, 14, 17, 18, 25, 40, 71, 74, 75 82, 83, 84, 95

Response rates were examined for potential weighting classes and these data were used to determine the final weighting classes. The potential classes were based primarily upon design characteristics, i.e., first stage sampling units and the four sample groups cited earlier for computation of the sampling weights. The weighting classes were defined as shown in Tables 1.10 and 1.11. With the exception of the first Washington weighting class, each weighting class for the initial telephone screenings consisted of three proximate area segments, or first stage sampling units. The area segments were first ordered in a serpentine manner by the Census geographic variables

State, County, Tract, TOTHUS,

as for the first stage sample selection, so that the geographic areas combined to form weighting classes were generally proximate to each other. For each site, the screening attempts for the No Previous Contact Sample were partitioned into two weighting classes of nearly equal size. The weighting classes were again formed from the ordered area segments. For each site, the screening attempts for Missed HU Sample formed a single weighting class. For Washington, the three segments with no Donnelley listings were also treated as a single weighting class.

After the weighting classes had been defined, adjusted sampling weights were computed within each weighting class. The weight adjustment procedure compensated for the fact that there were sample housing units (and/or Donnelley listings) for which the eligibility status was not known. For example, the eligibility status was not known for telephone attempts for which the final result was "Ring-No-Answer" and for field attempts where no contact was made.

Table 1.10. Response Distribution for Denver Screening Sample Weighting Classes

Weighting Class		Eligible, Responding	Eligible, Not Responding	Ineligible, Not Responding	Eligibility Not Known	Total
Sample	Segments					
Initial	01-03	56	18	2	6	82
Telephone	04-06	66	20	6	10	102
Screenings	07-09	53	19	7	9	88
	10-12	51	26	7	9	93
	13-15	65	25	6	9	105
	16-18	51	10	5	14	80
	19-21	68	23	7	8	106
	22-24	49	20	3	13	85
	25-27	67	10	4	8	89
	28-30	50	18	4	9	81
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	85-87	83	42	14	15	154
	88-90	63	19	1	13	96
	91-93	59	15	8	8	90
	94-96	71	20	3	3	97
	97-99	52	18	12	10	92
No Previous Contact Sample	01-53	55	14	20	44	133
	54-99	54	8	14	33	109
Missed HUs	01-99	27	8	0	13	48

Table 1.11. Response Distribution for Washington Screening Sample Weighting Classes

Weighting Class		Eligible, Responding	Eligible, Not Responding	Ineligible, Not Responding	Eligibility Not Known	Total
Sample	Segments					
Initial	001-004	58	15	11	21	105
Telephone	005-007	54	14	4	17	89
Screenings	008-010	69	9	5	11	94
	011-013	62	8	3	8	81
	014-016	47	12	2	15	76
	017-019	51	5	3	13	72
	020-022	48	7	5	7	67
	023-025	39	14	6	18	77
	026-028	32	8	15	10	65
	029-031	30	3	29	15	77
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	233-235	45	15	7	6	73
	236-238	37	13	8	17	75
	239-241	59	10	6	10	85
	242-244	50	12	4	22	88
	245-247	69	9	8	7	93
No Previous Contact Sample	001-106	41	17	26	83	167
	107-247	46	15	16	109	186
Missed HUs	001-250	16	3	1	26	46
No Donnelley Listings	248-250	60	16	1	19	96

An estimate of the proportion of eligible housing units in the h-th weighting class was computed as follows:

$$ELGPROP(h) = \Sigma\ WSCR(i,j)\ ELIG(i,j) / \Sigma\ WSCR(i,j)\ ELGKWN(i,j), \quad (15)$$

where each summation extends over all housing units that belong to weighting class h. The adjustment factor for housing unit nonresponse within the h-th weighting class was then computed as follows:

$$KHU(h) = \Sigma\ WSCR(i,j)\ ELGPROP(h) / \Sigma\ WSCR(i,j)\ RESP(i,j), \quad (16)$$

where each summation again extends over all housing units in the h-th weighting class. The numerator of this ratio estimates the number of eligible households that belong to weighting class h. This adjustment factor was used to produce the following adjusted weight:

$$WSCRADJ(i,j) = \begin{cases} WSCR(i,j)\ KHU(h) & \text{for responding households,} \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

for each housing unit that belongs to the h-th weighting class. The sum of the adjusted weights, WSCRADJ, over all housing units in the h-th weighting class is the estimated number of eligible housing units in the weighting class. This adjusted weight was used for analysis of the screening data as described in Chapter 2, both at the household level and at the person level since there was no person-level nonresponse within responding households for the screening interviews.

1.5 Selection of the Third Stage Sample

As discussed in Section 1.1, the eligible individuals for the CO study were the nonsmoking residents of the target areas between 18 and 70 years of age. The field screenings yielded 139 eligible individuals in responding households for the Denver samples, and 214 eligible individuals for the Washington samples. The field screenings were generated by the No Previous Contact Sample and the Missed HU Sample for Denver. They also

included the sample for the three area segments with no Donnelley listings for Washington. All eligible individuals from the Denver field samples were selected into the third stage sample for CO monitoring. For Washington, the 208 eligible individuals with a known telephone number were all selected for CO monitoring.

The initial telephone screenings yielded 2093 eligible individuals for Denver and 5209 eligible individuals for Washington. For each site, a stratified simple random sample of eligible individuals was selected for CO monitoring as shown in Tables 1.12 and 1.13. The purpose of the stratification was to oversample those individuals who appeared to be at greatest risk for high CO exposures based upon their screening data. Since the sample sizes defined by Tables 1.12 and 1.13 were not explicitly allocated to first-stage sampling units, the number of sample individuals selected from each first-stage unit is actually a random variable. Hence, the sample design is not strictly a nested design, which presents some problems for estimation of precision, as discussed in Section 1.7.

The sample stratification for the initial telephone screenings is shown in Tables 1.12 and 1.13. The strata are defined in terms of four stratification variables. Based upon discussions with EPA staff members it was decided that commuting time was the most important stratification variable. Respondents who had indicated a usual daily one-way commuting time of 30 minutes or more were defined to belong to the "long commuting time" strata. All other respondents were defined to belong to the "short commuting time" strata. It was decided that the second most important stratification variable was a gas applicance variable. This variable was defined to have the following three levels defined in terms of the gas

Table 1.12. Third Stage Sample Allocation for the Denver Sample

Stratum Number	Commuting Time	Gas Appliance	Attached Garage	Screened Eligibles	Sample Size
1	Short	Space Heater	Yes	48	48
2	Short	Space Heater	No	43	43
3	Short	Gas Stove[1]	Yes	70	56
4	Short	Gas Stove	No	342	148
5	Short	Other	Yes	529	229
6	Short	Other	No	577	250
7	Long	Space Heater or Gas Stove	-	143	74
8	Long	Other	Yes	182	78
9	Long	Other	No	159	74
Totals				2093	1000

[1]Gas stove, but not a space heater.

Table 1.13. Third Stage Sample Allocation for the Washington Sample

Stratum Number	Commuting Time	Gas Appliance	Attached Garage	Smoker in HU	Screened Eligibles	Sample Size	Lead Letter
1	Short	Space Heater	-	-	124	66	22
2	Short	Gas Stove[1]	Yes	-	162	66	22
3	Short	Gas Stove	No	Yes	447	143	48
4	Short	Gas Stove	No	No	1085	346	116
5	Short	Other	Yes	-	443	143	48
6	Short	Other	No	-	840	271	91
7	Long	Space Heater	-	-	94	66	22
8	Long	Gas Stove	Yes	-	105	66	22
9	Long	Gas Stove	No	Yes	295	94	31
10	Long	Gas Stove	No	No	730	233	78
11	Long	Other	Yes	-	286	94	32
12	Long	Other	No	-	598	191	64
Totals					5209	1779	596

[1]Gas stove, but not a space heater.

appliances used at the respondent's residence:

(a) a space heater is used,

(b) a gas stove is used, but not a space heater, or

(c) neither space heater nor gas stove is used.

The screening questionnaire had probed for use of several types of gas appliances in the residence. It was decided that all types of gas appliances other than space heaters and gas stoves were usually vented. Hence, space heaters and gas stoves were considered the major sources of CO generation within the home and were used for stratification. Rare groups based upon these two stratification variables were oversampled, and the remainder of the sample was proportionally allocated to the remaining strata for each site. Some of the larger strata that received a proportional allocation of the remaining sample were divided into substrata. Since the substrata received a proportional allocation, the additional stratification simply added control over distributional characteristics of the sample.

Two additional variables were used to define substrata: Presence of an attached garage and use of tobacco in the home. Individuals in households that indicated an attached garage were assigned to one level of the attached garage variable, and all other individuals were assigned to a second level. The individuals from a residence in which someone was identified as a tobacco user were assigned to one level of the smoking or tobacco-use variable. The individuals in all other households were assigned to a second level of this variable. The attached garage variable was considered to be the more important of these two substratification variables. The smoking variable was not used for the Denver stratification due to the smaller stratum sizes.

The third stage sample of persons depicted by Tables 1.12 and 1.13 was selected for each site as a stratified simple random sample without replacement. Hence, the person-level weight component was simple to compute. If N(k) represents the number of screened eligible individuals from stratum k, the person-level weight component was computed for each sample individual from stratum k as

$$W_8(k) = N(k) / n(k), \qquad (18)$$

where n(k) is the ultimate sample size from stratum k.

Another aspect of the third stage sample design is the allocation of individuals to specific days within the sample period. A major purpose of the CO study is to estimate the distribution of personal CO exposures for the study populations during the study season. Of particular interest is the maximum personal CO exposure. Individual CO exposure is heavily dependent upon two factors: weather and activity patterns. Since weather is such an important factor, it was necessary to field as large a sample as possible on each day during the study period. Otherwise, there could be no one monitored on the days with weather patterns producing the highest CO levels. Since activity patterns are important, the sample participants could not be allowed complete freedom of choice in selecting a day to be monitored. The sample subjects could introduce a bias by selecting mostly days when they plan to be inactive or stay at home. The strong influence of weather and activity patterns upon CO exposure suggests that a specific day should be randomly selected for each individal to be monitored. However, it was anticipated that the response rate would be so poor as to invalidate the study if only one day was offered for each sample subject to participate. Hence, the sample for Washington was randomly allocated to non-overlapping three-day interview periods. Likewise, the sample for

Denver was allocated to four-day interview periods. Four-day interview periods were used for the Denver sample because each individual selected for Denver was asked to participate for two consecutive days. Each individual in the Washington sample participated for only one day.

The allocation of individuals in the Washington sample to specific three-day interview periods had a greater negative impact on the response rate than had been expected. Some individuals indicated that they were willing to participate, but not within the selected time period. These individuals were given one additional opportunity to participate by randomly reallocating them to one new three-day period. A total of 550 individuals were reallocated in this manner for the Washington sample. Reallocation to new time periods was also allowed for the Denver sample. However, the method of reallocation was somewhat different.

The third stage sample for Washington also incorporated a lead letter methodology study. A sample of 596 individuals was selected to receive a lead letter. A random subsample was selected from each of the strata as shown in Table 1.13. The lead letter informed the individual that he or she had been selected for monitoring and that a telephone interviewer would be calling soon. The lead letters appear to have had a positive effect upon response rate. The overall response rate for individuals selected into the Washington sample was about 58 percent, but the response rate for individuals in the lead letter sample was approximately 63 percent.[1] Hence, a person-level response rate of about 65 percent may be possible for future studies of this type.

---

[1]These response rates are calculated as the number of individuals who agreed to schedule a monitoring appointment divided by the number of individuals selected.

1.6 Third Stage Response and Sample Weights

The initial, unadjusted sampling weights for persons selected for CO monitoring are the product of two weight components. The first component is the reciprocal of the probability of selection of the household for screening. The second is the reciprocal of the probability of selection of the person for monitoring, given that the household was selected. The product gives the reciprocal of the overall probability of selection of the person for monitoring. Hence, the initial person-level sampling weights were computed for Washington as follows:

$$\text{PIDWT0}(i,j) = \begin{cases} \text{WSCR}(i,j)\ W_8(k) & \text{for the j-th person in the i-th household and the k-th stratum for the initial telephone screenings,} \\ \text{WSCR}(i,j)\ (214/208) & \text{for the special field screenings.} \end{cases} \quad (19)$$

The first factor, WSCR, is the initial, unadjusted household-level sampling weight computed as shown in Section 1.4. The second factor for the individuals selected from the initial telephone screenings, W_8, is the person-level stratum weight component computed as shown in formula (18). The second factor for the special field screenings is 214/208 since the 208 eligible individuals with known telephone numbers were selected from the 214 eligible individuals screened in Washington, as discussed in Section 5.

Nonresponse occurs at two levels for the CO study design: at the household level for screening and at the person level for CO monitoring. The development of the screening weight, WSCRADJ, adjusted for household-level nonresponse was discussed in Section 1.4. Hence, the person-level weights adjusted for household-level nonresponse were computed for Washington as follows:

$$\text{PIDWT1}(i,j) = \begin{cases} \text{WSCRADJ}(i,j)\ W_8(k) & \text{for the j-th person in the i-th household and the k-th stratum for the initial telephone screenings,} \\ \text{WSCRADJ}(i,j)\ (214/208) & \text{for the special field screenings.} \end{cases} \quad (20)$$

The person-level weight factors for PIDWT1 are identical to those for the initial weight, PIDWT0, since this set of weights is only adjusted for household-level nonresponse.

Several types of data were collected for each participant in the CO study:

(1) Questionnaire data.
(2) CO monitor data.
(3) Activity diary data.
(4) Breath data.

Any one or all of these types of data could have been used to define a responding person for the CO study. It was decided, however, that most analyses would center on the CO monitor data. Hence, one set of person-level adjusted weights was developed treating the individuals with usable CO monitor data as the respondents. When a more restrictive condition applies for a given analysis, e.g., an analysis using only the persons with both monitor data and activity diary data, ratio adjustments within the person-level weighting classes can be used to adjust for the additional nonresponse. Alternatively, when a less restrictive condition applies, e.g., analysis of questionnaire data alone, ratio adjustment of the weights given by (20) within the person-level weighting classes will produce the appropriate analysis weights.

Weighting-class procedures were used to adjust for person-level nonresponse. For the initial telephone screenings, the sample selection strata shown in Table 1.13 were used as the weighting classes. For the sample of individuals from the special field screenings, the screener weighting classes shown in Table 1.11 were used as the person-level weighting classes as well, except that the Missed HU and No Donnelley Listing samples were combined into one weighting class as shown in Table 1.14. These two samples were combined because the missed housing units

Table 1.14. Person-Level Weighting Class for the Washington Sample

Weighting Class		Eligible,	Eligible,	Ineligible,	
Sample	Stratum	Responding	Not Responding	Not Responding	Total
Initial	1	18	48	0	66
Telephone	2	25	41	0	66
Screenings	3	41	101	1	143
	4	111	232	3	346
	5	71	71	1	143
	6	110	160	1	271
	7	23	42	1	66
	8	25	40	1	66
	9	40	54	0	94
	10	76	156	1	233
	11	40	54	0	94
	12	62	129	0	191
No Previous Contact Sample Segments 001 - 106		20	32	0	52
No Previous Contact Sample Segments 107 - 247		15	39	1	55
Missed HUs and No Donnelley Listings		35	65	1	101
Total		712	1264	11	1987

generated too few responding persons to be treated as a separate person-level weighting class. The individuals with usable CO data were treated as the respondents for the person-level weight adjustments. Due to problems with the CO monitors, approximately ten percent of the CO data collected were disregarded as invalid. The distribution of final result codes for all individuals selected for CO monitoring in Washington is shown in Table 1.15. It is easily seen from Table 1.15 that appointments were scheduled for about 58 percent of the individuals selected into the sample. However, due to various factors, usable CO monitor data were obtained for only about 36 percent of the individuals sampled. Instrument failure and refusal to carry the monitor were two of the major reasons for the low response rate.

The final weight adjustment factor for individuals with usable CO data was computed as follows:

$$\text{PIDADJ1}(h) = \Sigma_1 \text{ PIDWT1}(i,j) \; / \; \Sigma_2 \text{ PIDWT1}(i,j), \tag{21}$$

where Σ_1 represents summation over all eligible members of weighting class h, and Σ_2 represents summation over all responding members of weighting class h. Hence, the final analysis weights for the Washington sample were computed as follows for each member of weighting class h:

$$\text{PIDWT2}(i,j) = \begin{cases} \text{PIDWT1}(i,j)\ \text{PIDADJ}(h) & \text{if responding,} \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

The final analysis weights, PIDWT2, have been adjusted to reduce the potential bias due to both household-level and person-level nonresponse. These weight adjustment procedures reduce bias to the extent that respondents and nonrespondents are alike with respect to the probability of responding and/or the response values within weighting classes. As shown in Table 1.8, approximately 76 percent of the eligible households in the Washington screening sample responded. And, from Table 1.15 we see that

Table 1.15. Distribution of Final Result Codes for Individuals Selected for CO Monitoring in Washington

Result Code	Interpretation	Frequency N	Frequency %
02	No Contact after Appointment Scheduled	16	0.8
03	Need to be Rescheduled	1	0.1
04	Refused to Keep Appointment	40	2.0
12	Appointment not Kept - Will not Reschedule	76	3.8
14	Refused After Field Contact	51	2.6
15	Snowstorm Forced Cancellation	9	0.5
21	Completed Data Collection; Unusable CO Data	132	6.6
22	Partial Data Collected	7	0.4
23	Wrong Person Monitored	2	0.1
24	Monitor Malfunctioned	68	3.4
25	Other Result, Eligible for Monitoring	36	1.8
26	Ineligible (e.g., Smoking or Illiterate)	11	0.6
30	Usable CO Data	712	35.8
99	Unable to Schedule an Appointment	826	41.6
	Total	1987	100.1

usable CO monitoring data were obtained for about 36 percent of the individuals selected for monitoring in Washington. These response rates could be improved in the future by sending lead letters to all individuals selected for monitoring, by reduction of monitor failure and by making the monitors less intrusive. Like most personal monitoring studies, the CO study achieved a relatively low overall response rate. However, it may be very plausible to presume that the CO exposures of respondents and non-respondents are alike within weighting classes. If so, the low response rate is not as much of a problem as it might be in some other type of study, e.g., a study of people's attitudes and opinions.

1.7 Variance Estimation

The sampling design for the CO study is a deeply stratified three stage design. Area segments defined by 1980 Census block groups are selected at the first stage. Donnelley listings are selected at the second stage. However, the second-stage sample is a multi-phase sample. Each sample Donnelley listing is initially assigned to the telephone phase. A subsample of the listings not covered by the telephone phase, e.g., listings with no telephone number, is then selected for a second phase. In particular, the No Previous Contact Sample is a field interview phase of the second stage sample. Moreover, a third phase of the second stage sample is selection of a subsample of listings for the Missed HU Sample. Finally, the third stage sample of people for monitoring is not completely nested within the first-stage sampling units. As a result of these design complexities, exact formulas for estimation of variances and standard errors are complex, if not intractable. Approximate formulas are generally used to obtain appropriate estimates of standard errors with designs of this type.

For the CO study design a good method of approximation for standard errors would be a balanced half-sample pseudo-replication technique. With this technique the respondents would be randomly divided into two pseudo-replicates and the implicit sample selection strata from the first-phase design would be used to form super-strata. The variance estimation would then be essentially an estimate based upon differences between half-sample replicates within super-strata as discussed by McCarthy [1966 and 1969].

An alternative method for variance estimation has also been developed for the CO study design. This alternative is an estimate based upon the variability between first-stage sampling units (FSUs). The FSUs were selected from a deeply ordered sampling frame using a sequental, minimum probability replacement (MPR) sampling technique. The result is deep implicit stratification with one FSU selected from each ultimate stratum. The individual FSUs tended to have person-level sample sizes that were too small for them to be used directly for between FSU variance approximation. Hence, the ordering of the FSUs on the sampling frame was used first to combine proximate FSUs to form 43 analysis FSUs. The frame ordering was then used to combine FSUs into 21 analysis super-strata consisting of proximate pairs of analysis FSUs. Estimates of sample precision can then be generated easily using RTI's software for analysis of data from complex sample surveys. In particular, Taylor series approximations for the variability between analysis FSUs within super-strata are easily computed and yield reasonable approximations to the sampling variances for the CO study design.

It would be interesting to try both variance estimation techniques for some analysis statistics. Unfortunately, however, the limited budget for the CO study may preclude implementation of the more complex balanced half-sample technique. Since the CO study is a methodology study, its main

emphasis is not on analytic results. Methodological results are of primary concern, and it may be more appropriate to use the available resources to investigate methodological issues.

2. SCREENER ANALYSIS

Statistical analysis of the data from the screening sample consists mainly of estimating totals and proportions. Estimates are desired for both household-level and person-level data items in the Washington, D.C. and metropolitan Denver areas.

2.1 Theoretical Formulation

2.1.1 Estimating Item Totals

The form of the estimators for household-level and person-level data items are the same. The estimator associated with households simply sums over households with the required weighting to estimate the total number of households at a particular study site. Similarly, the estimator associated with estimating persons, is formulated as the weighted sum of individuals. Household and person-level weights within households are identical because no selection was done within households. The household total estimator for any data item within sites is as follows:

$$\hat{T}(HH) = \sum_i \sum_j WSCRADJ(i,j)Y(i,j),$$

where

$WSCRADJ(i,j)$ = the sampling weight adjusted for housing unit nonresponse corresponding to the j-th housing unit within the i-th first stage unit

$$Y(i,j) = \begin{cases} \text{the value for the data item of interest given the j-th housing unit within the i-th first stage unit for respondens} \\ 0, \text{ otherwise.} \end{cases}$$

The person-level total estimator, for each data item within sites, is expressed as

$$\hat{T}(PR) = \sum_i \sum_k WSCRADJ(i,k)Y(i,k),$$

where

WSCRADJ(i,k) = the sampling weight adjusted for housing unit nonresponse for the k-th individual within the i-th first stage unit

$$Y(h,i,k = \begin{cases} \text{the value for the data item of interest given the k-th individual within the i-th first stage unit for respondents} \\ 0, \text{ otherwise.} \end{cases}$$

2.1.2 Estimating Item Proportions

The proportions of concern are ratios of totals for specific categories (e.g., yes, no, etc.) to the overall estimated total. For any category of data item within site, the estimator of a proportion is formulated as

$$\hat{P}(I) = \frac{\hat{T}(CA)}{\hat{T}(OA)},$$

where

$\hat{T}(CA)$ = the estimated total for the specific category of the item of interest

$\hat{T}(OA)$ = the overall estimated total for the item of interest.

The estimator $\hat{P}(I)$, is used to estimate household and person-level proportions for both study sites.

2.1.3 The Associated Standard Errors

All standard errors are computed based on computer software which utilizes the delta method, i.e., Taylor series linearization variance approximation. This method is used frequently by various research agencies including the Bureau of the Census [1978]. It is utilized in the Current Population Survey to produce variance estimates for nonlinear statistics.

An important point to note is that, the Bureau of the Census uses the conventional variance formulation when weighting class adjustments have been incorporated. The CO study does the same. To the author's knowledge, all other research agencies use a similar variance approximation even when weighting class adjustments have been implemented. The conventional method treats weighting class adjustments, or post-stratification adjustments, as constants in the Taylor series variance formulation. Extensive research (Jones and Chromy [1982]) indicates that when the nonresponse rate is as low as 10%, the conventional approximation may vastly overestimate, or underestimate the "true" sampling variance. Conversely, the Jones-Chromy variance estimators yield estimates which are consistently less biased even at a high (50%) nonresponse rate. Their work determined that the direction of nonresponse bias depends primarily on

a. severity of nonresponse,

b. type of weighting class adjustment(s) used, and

c. amount of variation among data.

Hence, standard error estimates presented for the CO study are thought to be inflated for the following reasons:

a. a contribution due to unequal selection probabilities,

b. a contribution due to estimating the proportion of eligible housing units in the h-th weighting class, and

c. a contribution resulting from estimating the actual weighting class adjustment factor, that is, the ratio of the two functions:

    1. an estimator for the total number of <u>eligibles</u> in the h-th weighting class

    2. an estimator for the total number of <u>responding</u> eligibles in the h-th weighting class

The conventional Taylor approximation takes into account only the contribution made from the sampling variance, however, the Jones-Chromy variance

estimators are formulated to handle the sampling variance and weighting class adjustment contributions. The estimators can be modified to accommodate the contribution from estimating the proportion of eligible housing units.

There is hope that progress will be made toward modifying inhouse statistical software to handle all of the above effects.

2.2 Results

2.2.1 Household Screener Statistical Analysis

Gas household appliances and certain sources of home heating are considered producers of carbon monoxide. The following related questions of interest were asked of an eligible respondent, and analyzed for each household:

Is there

a. a fire place which is used?

b. a wood stove?

c. a gas furnace?

d. a gas cooking stove?

e. a gas hot water heater?

f. a gas clothes dryer?

g. a gas or kerosene space heater?

h. any other gas appliances?

i. an attached garage, or share a multi-family garage?

Estimates from these items are presented in Tables 2.1-2.8. These results are based on successful interviews from 4394 households in the Washington, D.C. metropolitan area, and 2128 households from the Denver metropolitan area. Sample sizes are constant for both sites over all

Table 2.1. Estimated Number of Households Using a Fireplace

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	316,925	0.3323	103,211	0.2991
Standard Error	29,538	0.0301	16,030	0.0362
No	636,322	0.6672	241,629	0.7002
Standard Error	59,462	0.0302	18,608	0.0361
Not Known	467	0.0004	227	0.0006
Standard Error	193	0.0002	141	0.0004
Grand Total	953,714		345,068	
Sample Size	4,394		2,127	

Table 2.2. Estimated Number of Households Using a Wood Stove

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	37,721	0.0395	20,314	0.0588
Standard Error	6,858	0.0075	2,841	0.0089
No	915,382	0.9598	324,453	0.9400
Standard Error	64,623	0.0075	24,594	0.0089
Not Known	611	0.0006	396	0.0011
Standard Error	378	0.0003	184	0.0005
Grand Total	953,714		345,163	
Sample Size	4,394		2,128	

Table 2.3. Estimated Number of Households Using a Gas Furnace

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	532,347	0.5581	245,902	0.7124
Standard Error	40,943	0.0355	22,231	0.0379
No	391,130	0.4101	94,964	0.2751
Standard Error	52,169	0.0365	14,608	0.0376
Not Known	30,237	0.0317	4,297	0.0124
Standard Error	13,517	0.0146	1,246	0.0036
Grand Total	953,714		345,163	
Sample Size	4,394		2,128	

Table 2.4. Estimated Number of Households Using a Gas or Kerosene Space Heater

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	30,530	0.0320	18,352	0.0531
Standard Error	5,621	0.0050	4,480	0.0111
No	921,536	0.9662	326,152	0.9449
Standard Error	61,817	0.0050	22,301	0.0111
Not Known	1,649	0.0017	659	0.0019
Standard Error	444	0.0004	265	0.0007
Grand Total	953,714		345,163	
Sample Size	4,394		2,128	

Table 2.5. Estimated Number of Households Using a Gas Cooking Stove

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	609,029	0.6385	85,542	0.2478
Standard Error	49,353	0.0329	7,308	0.0245
No	344,329	0.3610	259,365	0.7514
Standard Error	41,033	0.0329	24,195	0.0245
Not Known	357	0.0003	256	0.0007
Standard Error	213	0.0002	160	0.0004
Grand Total	953,714		345,068	
Sample Size	4,394		2,128	

Table 2.6. Estimated Number of Households Using a Gas Hot Water Heater

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	542,855	0.5692	269,810	0.7816
Standard Error	37,564	0.0308	20,930	0.0332
No	361,036	0.3785	67,425	0.1953
Standard Error	50,113	0.0346	13,256	0.0331
Not Known	49,823	0.0522	7,928	0.0229
Standard Error	13,628	0.0152	1,175	0.0037
Grand Total	953,714		345,163	
Sample Size	4,394		2,128	

Table 2.7. Estimated Number of Households Using a Gas Clothes Dryer

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	191,803	0.2011	57,402	0.1663
Standard Error	17,316	0.0195	8,018	0.0179
No	737,472	0.7733	285,830	0.8281
Standard Error	64,529	0.0217	20,381	0.0178
Not Known	24,284	0.0254	1,931	0.0055
Standard Error	13,259	0.0142	724	0.0021
Grand Total	953,560		345,163	
Sample Size	4,393		2,128	

Table 2.8. Estimated Number of Households Using Other Gas Appliances

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	16,036	0.0168	782	0.0022
Standard Error	3,362	0.0037	293	0.0008
No	934,713	0.9800	344,019	0.9966
Standard Error	64,821	0.0038	24,659	0.0009
Not Known	2,966	0.0031	362	0.0010
Standard Error	704	0.0007	172	0.0005
Grand Total	953,714		345,163	
Sample Size	4,394		2,128	

data items, with the exception of isolated missing item data for which the sample sizes and population estimates are reduced accordingly. The population estimate for the number of households in the two sites for the time the survey was conducted is 953,714 for Washington and 345,163 for Denver. Due to the fact that no item-level nonresponse adjustments or item imputations were made, these estimates are underestimates of the true total. However, the difference between the estimates and the true totals are believed not to be statistically significant.

For each data item, a separate category (namely, "not known") was created to represent those households for which the respondent or interviewer was uncertain whether a particular appliance, etc. existed. As an example, the study determined that there are an estimated 467 households in Washington, D.C. (Table 2.1) for which the screening questionnaire cannot ascertained whether a fireplace is being used.

In terms of home heating, about 33 percent of the Washington area households have fireplaces in use and roughly 67 percent do not (Table 2.1). Fewer than 30 percent of Denver households have fireplaces and about 70 percent do not have fireplaces (Table 2.1).

As expected, there are fewer homes utilizing wood stoves - an estimated 4 percent for the Washington area and 6 percent for Denver (Table 2.2). Among the four means of home heating, the gas furnace is clearly the most common heating source. This is indicated in Table 2.3: 56 percent or 532,347 households use gas furnaces in the Washington area and 71 percent or 245,902 households use gas furnaces in the Denver area. The percent of households using gas or kerosene space heaters is similar to those using wood stoves, about 3 percent for Washington and 5 percent for Denver (Table 2.4).

Statistical results for the usage of gas appliances are presented in Tables 2.5-2.8. It is estimated that over half of Washington, D.C. households utilize gas cooking stoves. On the other hand, about a quarter of Denver households use gas cooking stoves. This difference may be due to a possible difference in utility rates in the two metropolitan areas. In particular, an estimated 64 percent or 609,029 households in Washington, D.C., and 25 percent or 85,542 households in Denver use gas cooking stoves (Table 2.5). Gas hot water heaters serve 542,855 or 57 percent and 269,810 or 78 percent of households from the respective sites, Washington and Denver (Table 2.6). An estimated 191,803 households or 20 percent and 57,402 households or 17 percent use gas clothers dryers in their homes (Table 2.7). Other gas appliances are seldom used. This is evident in Table 2.8: approximately 2 percent and 0.2 percent of Washington and Denver metropolitan area households use other gas appliances, respectively.

According to Table 2.9, a combined estimate of 207,719 or 22 percent of Washington area households have an attached garage or share a multi-family garage. This compares to a combined estimate of 120,460 or 35 percent for Denver area households.

2.2.2 Personal Item Statistical Analysis

As with most studies which involve a sample of households, there is some interest in specific attributes of the household members. In the CO Study an eligible respondent was requested to answer several questions on behalf of each individual living in the household. Below is the information asked for each individual:

a. Sex

b. Age

c. Individual's relationship to head of household

Table 2.9. Estimated Number of Households Having an Attached Garage, or Share a Multi-Family Garage

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Attached Garage	165,919	0.1740	108,934	0.3156
Standard Error	22,525	0.0191	11,188	0.0284
Multi-Family Garage	41,800	0.0438	11,526	0.0333
Standard Error	13,845	0.0153	3,919	0.0099
Neither	744,976	0.7812	222,199	0.6437
Standard Error	58,749	0.0250	18,870	0.0298
Not Known	862	0.0009	2,504	0.0072
Standard Error	426	0.0004	2,469	0.0070
Grand Total	953,557		345,163	
Sample Size	4,393		2,128	

d. Whether individual presently smokes or uses tobacco in any form

e. Whether individual is employed either full or part time

f. Whether individual travels to and from work, school, or any other place at least 3 times a week

g. Amount of time spent traveling one way when going any place at least 3 times per week

Statistical results to follow are based on varying sample sizes due entirely to missing item data. However, it is worth noting that sample sizes for the sex question are 11,545 for Washington D.C. and 5,142 for Denver. Data for sex could be determined in a couple of ways, so other samples sizes which correspond to other data items, are simply subsets of the sample of sex data. Similar to the household-level data, no item-level nonresponse adjustments or imputations were made. Thus, personal item estimates are considered underestimates of the true total. The nonresponse bias should have little effect upon proportion estimates.

The category "not known" is presented to represent those individuals for whom the respondent or interviewer could not determine a correct entry for any individual's attributes. Included in this category is a population estimate for the data item of interest. The statistical results of all personal item data are presented in Tables 2.10-2.16.

According to Table 2.10, an estimated 1.29 million people or 48 percent of the Washington metropolitan population are male. This compares to the 1980 Census which reports 1.33 million males or 48 percent in the same area (Table 2.11). Also, 1.38 million females or 52 percent are estimated from the CO Study results versus 1.44 million females or 52 percent from the 1980 Census. Neither of these statistics are expected to be exactly equal, primarily because of a population change from 1980 to 1983 and sampling variability. They are however in the same statistical realm.

Table 2.10. Estimated Sex Distribution

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Male	1,286,056	0.4821	415,730	0.4739
Standard Error	133,431	0.0094	30,762	0.0091
Female	1,379,812	0.5173	460,836	0.5253
Standard Error	116,697	0.0094	39,728	0.0091
Not Known	1,329	0.0004	556	0.0006
Standard Error	483	0.0001	402	0.0004
Grand Total	2,667,197		877,122	
Sample Size	11,545		5,142	

Table 2.11. The Sex Distribution According to the 1980 Census

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Male	1,327,797	0.4805	380,479	0.4867
Female	1,435,308	0.5195	401,305	0.5133
Population Total	2,763,105		781,784	

When sex is disregarded, the CO Study results estimate the 1983 Washington metropolitan population to be 2.67 million versus 2.76 million for the 1980 Census. Denver metropolitan statistics by sex can be seen in Tables 2.10 and 2.11. The proportions are quite comparable; CO study estimates are 47 percent males and 53 percent females compared to 1980 Census statistics of 49 percent males and 51 percent females.

Ages have been grouped for the sake of complying with Census group definitions. These age groups are:

(1) under 18 years of age
(2) between 18 and 64 years of age and
(3) age 65 and older.

Estimated age group totals and proportions for both sites, compare quite well with Census statistics. Also, the age distributions for the two sites are nearly identical. CO Study estimates for the Age Groups are presented in Table 2.12 and corresponding 1980 census results are in Table 2.13. As can be seen the percent of individuals under 18 years, between 18 and 64 years, are respectively, about 26 percent and 65 percent (ignoring site and reference source). The comparison for persons 65 years and older is not quite as good.

Census data are not available for other items of interest. The CO Study results for relationships to head of household screened are categorized in Table 2.14. As anticipated, more head of households were screened than any other category, 35 percent and 40 percent for Washington and Denver metropolitan areas.

For the remaining person-level data items, statistical analysis was conducted only for those individuals age 13 and older. Also, estimated totals from data items to follow are far more affected by bias due to item nonresponse than previous data items. The nonresponse is indicated by a

Table 2.12. Estimated Age Distribution - Categorized According to the 1980 Census Definitions

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Under 18 Years	699,556	0.2677	227,112	0.2626
Standard Error	108,889	0.0197	27,009	0.0155
Between 18 to 64 Years	1,720,330	0.6583	564,554	0.6528
Standard Error	139,452	0.0150	46,137	0.0142
65 Years and Older	153,628	0.0587	67,091	0.0775
Standard Error	10,816	0.0055	5,959	0.0092
Not Known	39,405	0.0150	5,939	0.0068
Standard Error	7,320	0.0024	822	0.0010
Grand Total	2,612,919		864,695	
Sample Size	11,188		5,015	

Table 2.13. Age Distribution According to the 1980 Census

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Under 18 Years	721,170[1]	0.261	197,171[2]	0.2522[2]
Between 18 to 64 Years	1,826,412[1]	0.661	506,955[2]	0.6484[2]
65 Years and Older	215,523[1]	0.077	77,658[2]	0.0994[2]
Population Total	2,763,105		781,784[2]	

[1]Estimated from proportions published in the 1980 General Population Characteristics for District of Columbia.

[2]Estimated from city proportions published in the 1980 General Population Characteristics for Colorado.

Table 2.14. Estimated Distribution of Relationship to Head of Household

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Head of Household	949,796	0.3549	346,637	0.3953
Standard Error	63,594	0.0126	24,712	0.0099
Spouse	554,922	0.2073	191,165	0.2180
Standard Error	49,274	0.0180	14,866	0.0105
Child of Head	891,611	0.3332	229,522	0.2617
Standard Error	151,940	0.0293	16,388	0.0144
Other Relation	126,729	0.0473	30,882	0.0352
Standard Error	20,789	0.0088	3,969	0.0053
No Relation	147,567	0.0551	41,584	0.0474
Standard Error	26,472	0.0078	8,666	0.0087
Not Known	5,069	0.0018	36,991	0.0421
Standard Error	2,533	0.0009	21,100	0.0214
Grand Total	2,675,694			
Sample Size	11,543		5,138	

reduction in sample size. The point should be emphasized that these proportions are probably reliable, since proportions are not affected as much by nonresponse bias (i.e. if the likelihood of obtaining a response is a random variable). The Washington and Denver study site findings indicate an estimated 33 percent and 38 percent of the respective populations smoke or use tobacco in some form (Table 2.15). Between 70 percent and 72 percent of the individuals in these sites work either full or part time (Table 2.16). Persons who travel anywhere at least 3 times per week are estimated to be between 82 percent and 84 percent (Table 2.17) for the study sites. The majority of persons travelling (anywhere at least 3 times per week) are estimated to take short trips, actually taking less than 15 minutes (Table 2.18). This is true for both sites.

Table 2.15. Estimated Distribution of Persons 13 Years and Older Who Smoke or Use Tobacco in Any Form

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	639,739	0.3318	244,884	0.3795
Standard Error	69,002	0.0180	29,266	0.0234
No	1,287,619	0.6678	400,079	0.6200
Standard Error	85,016	0.0180	23,164	0.0234
Not Known	633	0.0003	254	0.0003
Standard Error	324	0.0001	124	0.0001
Grand Total	1,927,991		645,217	
Sample Size	8,791		3,953	

Table 2.16. Estimated Distribution of Persons 13 Years or Older Who Work Either Full or Part Time

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	1,333,061	0.7018	464,960	0.7207
Standard Error	96,555	0.0140	38,133	0.0159
No	561,831	0.2958	179,528	0.2783
Standard Error	42,112	0.0139	13,535	0.0158
Not Known	4,435	0.0023	602	0.0009
Standard Error	1,055	0.0005	258	0.0004
Grand Total	1,899,327		645,090	
Sample Size	8,778		3,951	

Table 2.17. Estimated Distribution of Persons 13 Years or Older Who Travel Anywhere at Least 3 Times Per Week

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Yes	1,606,757	0.8406	530,420	0.8221
Standard Error	110,928	0.0085	41,049	0.0130
No	297,010	0.1553	110,533	0.01713
Standard Error	28,404	0.0084	9,833	0.0127
Not Known	7,639	0.0039	4,189	0.0064
Standard Error	2,785	0.0015	2,571	0.0039
Grand Total	1,911,406		645,143	
Sample Size	8,787		3,952	

Table 2.18. Estimated Distribution of Amount of Time Spent Traveling One Way At Least 3 Times Per Week for Persons 13 Years or Older

	Washington, D.C. Metropolitan Area		Denver Metropolitan Area	
	Total	Proportion	Total	Proportion
Less than 15 minutes	1,002,041	0.4748	357,480	0.5226
Standard Error	84,829	0.0119	20,647	0.0239
16-30 Minutes	686,847	0.3254	257,484	0.3764
Standard Error	66,360	0.0121	29,920	0.0228
31-45 Minutes	216,547	0.1026	28,013	0.0409
Standard Error	21,000	0.0056	4,445	0.0066
Above 45 Minutes	186,342	0.0883	33,724	0.0493
Standard Error	22,119	0.0067	7,775	0.0097
Not Known	18,380	0.0087	7,314	0.0106
Standard Error	2,015	0.0011	3,433	0.0051
Grand Total	2,110,157		684,014	
Sample Size	9,268		4,152	

3. METHODOLOGICAL RESULTS

3.1 Introduction

The sample design for the CO monitoring project incorporated three methodology studies:

(1) The use of telephone directory listings classified by Census geographic variables in association with standard area household sampling techniques to identify sample housing units.

(2) The use of a lead letter mailed to sample subjects stating that they will be called to schedule an appointment for monitoring.

(3) Sampling person-days for monitoring, rather than simply selecting persons and letting each person choose a day to be monitored.

The purpose of this chapter is to report the results of these methodological studies.

3.2 Use of Geographically Classified Telephone Directory Listings in Association with Standard Area Household Sampling Techniques

As discussed in Chapter 1, the sample design for the CO study was a deeply stratified, three stage design. The EPA purposively selected the metropolitan areas surrounding Washington, D.C., and Denver, Colorado, as the study sites. Area sample segments defined by Census block groups were selected at the first stage of sampling. A computer tape listing the selected block groups was then sent to Donnelley Marketing Corporation. The tape was returned with computerized listings of names, addresses, and telephone numbers for the selected block groups. A sample of listings was then selected for each first stage sampling unit to identify the sample housing units. The screening interviews were conducted by telephone for all sample listings with a telephone number. Since the target population included households other than those with listed telephone numbers, field screening interviews were also necessary. A subsample of the listings for which a telephone interview was not possible (see Table 1.1) was selected

for field screening. Also, a subsample of listings was selected for a missed housing unit (missed HU) check.

The missed HU subsample consisted of 150 listings for Denver and 300 listings for Washington. In each case, the FSUs with the number of 1980 Census occupied housing units 50 percent or more greater than the number of commercial listings were deliberately oversampled as shown in Table 1.3. The missed HU check was implemented by using standard field listing protocol to produce a unique geographic ordering on Census maps for each FSU. Each listing in the missed HU sample was located in the field. The interviewer then proceeded to the next housing unit as identified by the geographic ordering and checked to see if that housing unit was on the commercial list for the FSU. If not, a screening interview was attempted and the check was continued at the next housing unit. When the next housing unit was found to be on the commercial list, the missed HU check was complete. Technically, a screening interview should not have been conducted if the missed HU was on the complete frame of commercial listings and was simply missclassified with regard to Census block group. Interviews were conducted for all missed HUs, regarding them as not simply missclassified, partly to check the completeness of the listings. In most cases, missed HUs occurred in groups of one or two. In one instance, an entire block face of five HUs was missed (See Table 3.1). These five missed HUs were regarded as missclassified, and their data were disregarded for analyses and selection of participants for CO monitoring.

The results of the missed housing unit checks for Denver and Washington are summarized in Table 3.1. For each study site, approximately two percent of the listings were found to not belong to the FSU, or area seg-

Table 3.1 Results of Missed HU Checks

Result	Denver[1] N/%	Washington[2] N/%
A. Start Address Inside Area Segment		
1. Completed missed HU check and found no missed HUs.	108/72.0	203/67.7
2. Completed missed HU check and found one or more missed HUs	18/12.0[3,4]	22/7.3
3. Invalid start address	8/5.3	2/0.7
4. Could not locate start address due to incomplete Donnelley listing	0/0.0	1/0.3
5. Could not identify the apartment at which to begin the missed HU check	12/8.0	39/13.0
6. Found one or more missed HUs but not able to complete missed HU check (unable to match names in the Donnelley listings to apartment numbers)	1/0.7	4/1.3
7. Start address and next address were both office complexes; missed HU check was aborted	0/0.0	8/2.7
8. Start address was inside an old age or convalescent home; missed HU check aborted as group quarter were ineligible	0/0.0	16/5.3
B. Start Address Outside Area Segment		
1. Completed missed HU check and found no missed HUs	0/0.0	2/0.7
2. Completed missed HU check and found exactly one missed HU	0/0.0	1/0.3
3. Could not identify apartment at which to begin the missed HU check	1/0.7	0/0.0
4. Aborted missed HU check after traveling one mile, to first corner, or listing nine missed HUs	2/1.3[4]	2/0.7[4]
Total	150/100.0	300/100.0

[1]Field work done by Research Triangle Institute.

[2]Field work done by PEDCo Environmental, Inc. under a separate contract.

[3]In one case an entire block face of five HUs was missed. The data for these five HUs was disregarded. These missed HUs were regarded as missclassified.

[4]The data for these missed HUs was disregarded. These missed HUs were regarded as missclassified.

ment, to which they had been classified by Donnelley Corporation. Although a unique geographic ordering was not possible for listings outside the assigned area segment, a missed HU check was attempted for these listings. The purpose of this check was mainly to investigate the completeness of the commercial listings. The results in Table 3.1 for start addresses outside the segment would seem to indicate that clusters of HUs, e.g., block faces, tend to be misclassified occassionally and that random missclassification of individual HUs also occurrs.

When the missed HU start address was inside an apartment complex, implementation of the missed HU check was sometimes quite difficult. Table 3.1 shows that it was not possible to complete the missed HU check in apartment complexes for about 15 percent of the listings selected for the missed HU check. Since the listings did not generally include apartment numbers, it was necessary to get apartment numbers from mail boxes, apartment managers, or apartment residents. Sometimes these sources proved fruitless. Many of the instances in which a missed HU check could not be begun occurred in restricted-access apartments. Missed HUs seemed to occur more frequently in apartment complexes than in other areas when the check could be implemented. This may be due to the more transient nature of apartment dwellers. Only one missed HU check identified an entire block face (of five HUs) that had been missed by the commercial listings within the selected area segments. The general impression was that the commercial listings provided a reasonably complete listing of housing units.

Most of the listings were found to be correctly classified according to block group. Donnelley Corporation claims that their listings are about 95 percent complete. Our experience is not inconsistent with this claim. However, we found that the undercoverage does not seem to occur at random.

Instead, there were small geographic areas for which there were no listings whatsoever. The telephone directories for these areas simply had not been used in compiling the listings. Standard field procedures were used to list all housing units and select clusters of sample housing units for these block groups.

The major problem encountered in using the listings to identify sample housing units was that it was often difficult to locate the housing units corresponding to the sample listings in the field. The addresses generally came from telephone directory listings. Hence, most residents of apartment complexes all had the same address, namely the street address of the apartment complex. This presented some problem for location of the sample housing units. But, more importantly, it made the check for missed housing units very difficult to implement correctly. Because of these problems, and other more subtle problems with the operational definition of missed housing units, there seems to be no completely satisfactory way to perform the check for missed housing units for a sample from the commercial listings.

Based upon RTI's cited experience using geographically-classified telephone directory listings, it appears that such listings can best be utilized with a dual frame sampling procedure. In the dual frame approach, two independent samples are selected from the (complete) area frame:

(1) One sample is a standard area sample with sample clusters identified from field listings of all housing units in the selected area segments.

(2) The other sample uses the commercial listings to identify sample clusters in the selected area segments.

It is recommended that the commercial listing sample be used only to generate telephone interviews based upon the telephone directory listings. Using this methodology, the standard area frame sample is used to compen-

sate for the bias resulting from the telephone interviews generated by the commercial listing sample. In order to compensate for this bias, it is necessary to determine whether or not each household in the standard area frame sample is included on the commercial listing frame. This is easily done for commercial listings that come directly from the current telephone directory. A single questionnaire item can determine whether or not the household is served by a residential telephone number that is listed in the current telephone directory. It is not so easy to determine telephone coverage with respect to commercial listings based upon vehicle registrations. Hence, it is recommended that the vehicle registration records be disregarded. See Whitmore, et. al. [1983] for further discussion of these design recommendations.

Use of only the telephone directory listings for the commercial listing sample makes implementation of the dual frame methodology very straightforward. States with and without vehicle registration records in the commercial listings are handled in exactly the same way. For every sample household, one or two questionnaire items can be used to determine the number of residential telephone numbers that are listed in the current telephone directory for the household. This information is sufficient to facilitate unbiased estimation for linear statistics using either multiple frame multiplicity estimators, such as those discussed by Casady and Sirken [1980], or difference estimators, such as those discussed by Konijn [1973]. The difference estimators may be preferable since they address the bias correction more directly.

The CO study found that there were some area segments with no commercial listings whatsoever. Hence, a determination of whether or not telephone directory listings are available is needed for each area segment in the standard area household sample. If telephone directory listings are

not available for some area segments in the standard area household sample, the households in these area segments must be treated for estimation as not represented in the frame of telephone directory listings. Otherwise, all households with a currently listed telephone number are treated as being present on the frame of telephone directory listings.

For monitoring studies, such as the CO study, it is recommended that this dual frame approach be implemented as two half samples to generate the screening interviews. This will generally be less costly than obtaining the same number of interviews from a standard area frame sample alone. Some savings will be achieved by using the commercial lists instead of lists of housing units produced by field staff to identfy sample clusters. The use of telephone interviews instead of field interviews may produce some additional savings.

This dual frame approach could, of course, also be used for a field half-sample and a random-digit-dial (RDD) telephone half-sample. Some advantages of using the geographically-classified telephone directory listings instead of random digit dialing for one half sample are the following:

(1) Census geographic variables can be used to oversample subpopulations of interest at the first stage.

(2) The proportion of telephone numbers called that are working residential numbers will be much higher for the telephone directory sample.

(3) When field follow-up interviews are necessary, such as for personal monitoring studies, the geographic clustering will reduce subsequent field interview costs.

Of course, there is some loss in precision due to clustering and due to use of the incomplete telephone directory frame. These losses will generally be compensated by decreased cost for the sample survey. Thus, the proposed design is expected to be cost effective for monitoring studies.

3.3 Lead Letter Results

The sample design for Washington incorporated a lead letter methodology study, as described in Chapter 1. A random sample of 596 of the individuals selected for monitoring was sent a lead letter. The purpose of this lead letter was to inform the sample subject that he or she had been selected for monitoring and that an interviewer would be calling to schedule an appointment for monitoring. The individual was thanked for participating and the importance of the study was stressed. The lead letters appear to have had a positive effect upon response rate. The overall response rate for individuals selected into the Washington sample was about 58 percent, but the response rate for individuals in the lead letter sample was approximately 63 percent.[1] Hence, a person-level response rate of about 65 percent may be possible for future studies of this type using lead letters for all individuals selected for monitoring.

3.4 Sampling Person-Days

Individual exposure to carbon monoxide is heavily dependent upon both weather patterns and personal activity patterns, as discussed in Chapter 1. Since weather is such an important factor, it was necessary to field as large a sample as possible on each day during the study period. Otherwise, there could be no one monitored on the days with weather patterns producing the highest CO levels. Since personal activity patterns are important, the sample participants could not be allowed complete freedom of choice in selecting a day to be monitored. The sample subjects could introduce a bias by selecting mostly days when they plan to be inactive or stay at

---

[1]These response rates are calculated as the number of individuals who agreed to schedule a monitoring appointment divided by the number of individuals selected.

home. The strong influence of both weather patterns and personal activity patterns upon CO exposure suggests that a specific day should be randomly selected for each individual to be monitored. However, such a procedure is totally impractical. The response rate would surely be so poor as to invalidate the study if only one day was offered for each sample subject to participate. Hence, the sample for Washington was randomly allocated to non-overlapping three-day interview periods. This procedure had a greater negative impact on the response rate than was at first anticipated. Some sample members indicated that they were willing to participate, but not within the selected time period. These individuals were given one additional opportunity to participate by randomly reallocating them to one new three-day period. A total of 550 individuals were reallocated in this manner for the Washington sample.

The methodology for person-day sampling that was used for the Washington CO study is somewhat awkward. The reassignment to a new 3-day interview period required constant interaction between field staff and sampling staff during data collection. It also required continual updating of sampling files. Hence, the sampling task was much more expensive than that of a typical sample survey. A better procedure needs to be found for future air monitoring studies. The procedure must control allocation of persons to days and still be somewhat flexible with regard to the allocation to days. A methodology study to explore alternative methods of person-day sampling for studies monitoring personal exposure to airborne pollutants may be needed.

One methodology for person-day sampling could be the following. Suppose six days were randomly selected within the study period for each person selected for monitoring. These could be three consecutive days in

one week and the same three days in the next week. Suppose further that priorities from one to six were assigned to the days selected for each person. Each person selected for monitoring could then be told something like the following:

> You recently participated in an EPA-sponsored study by responding to a short questionnaire. You were told that someone might be calling on you to participate in a personal monitoring follow-up study. You have been selected for participation in the follow-up study. Due to the influence on exposure by weather patterns and personal activity patterns, it is necessary to monitor a representative sample of people of each day in the study period. Hence, I have an ordered list of seven days that have been selected for you. You are asked to choose the first of these days on which you can possibly participate. Otherwise, your data may be discounted in the analyses and not have as much impact as it would if you had participated on the first available day. Hence, would it be possible for your to participate on <Day 1>?

If the person could not participate on Day 1, the interview would proceed to Day 2, etc. A short reminder that the person needs to choose the first day on which participation is possible might be appropriate between offering Day 1 and Day 2.

Some type of weight adjustment procedure could then be used to compensate for the bias due to self-selection of days. For example, a weight standardization could be performed using a covariance model. This type of procedure could be used to adjust each day's sample to a standardized population, based, for example, on age, race, sex, and occupation. This might require a fairly large sample, e.g., enough to produce more than 25 respondents, on each day.

3.5 Conclusions

Important new information was learned from each of the three methodology studies of the CO project. It was found that geographically classified telephone directory listings can be used in a cost-effective manner in association with standard area household sampling techniques for personal

monitoring studies like the CO study. The sampling design for the cost-effective use of thse telephone directory listings differs radically from the design used for the CO study, however. Sending lead letters to individuals who were selected for personal monitoring prior to calling to schedule an appointment was found to be an effective strategy. The need for person-day sampling for studies that monitor personal exposure to airborne pollutants is apparent. The CO study gained valuable experience with this technique. Further study, possibly even another methodology study, is needed to refine this technique.

REFERENCES

1. Casady, Robert J. and Sirken, Monroe G. [1980]. A Multiplicity Estimator for Multiple Frame Sampling. Proceedings of the American Statistical Association Section on Survey Research Methods, 601-605.

2. Chromy, James R. [1979]. Sequential Sample Selection Methods. Proceedings of the American Statistical Association Section on Survey Research Methods, 401-406.

3. Cochran, W. G. [1963]. Sampling Techniques, 2nd ed., John Wiley and Sons, New York, pp. 327-353.

4. Jones, S. M. and J. R. Chromy [1982]. Improve Variance Estimators Using Weighting Class Adjustments for Sample Survey Nonresponse. American Statistical Association 1982 Proceedings of the Section on Survey Research Methods, pp. 105-110.

5. Konijn, H. S. [1973]. Statistical Theory of Sample Survey Design and Analysis. American Elsevier Publishing Company, New York, 126-132.

6. McCarthy, Philip J. [1966]. Replication: An Approach the the Analysis of Data from Complex Surveys (NCHS Vital and Health Statistics Series 2 Number 14). Public Health Service, Washington, D.C.

7. McCarthy, Philip J. [1969]. Pseudoreplication: Further Evaluation and Application of the Balanced Half-Sample Technique (Vital and Health Statistics Series 2 Number 31). Public Health Service, Washington, D.C.

8. Raj, Des [1968]. Sampling Theory. McGraw-Hill, New York, pp. 139-163.

9. U.S. Department of Commerce, Bureau of the Census. [1978]. The Current Population Survey: Design and Methodology. Technical Paper No. 40. U.S. Government Printing Office, Washington, D.C.

10. Williams, Rick L. and Chromy, James R. [1980]. SAS Sample Selection Macros. Proceedings of the Fifth Annual SAS Users Group International Conference.

11. Whitmore, R. W., Mason, R. E., Hartwell, T. D., and M. S. Rosenzweig [1983]. Use of Geographically Classified Telephone Directory Listings in Multi-Mode Surveys. To appear in: Proceedings of the American Statistical Association, Survey Research Methods Section